Two-component signalling systems of chloroplasts: function, distribution and evolution

Puthiyaveetil, Sujith

Two-component signalling systems of chloroplasts: function, distribution, and evolution

Sujith Puthiyaveetil

School of Biological and Chemical sciences

Queen Mary, University of London

A thesis submitted for the degree of Doctor of Philosophy

Statement of originality

This thesis incorporates results of experiments undertaken in collaboration with Prof. Tony A Kavanagh and Dr. Christine Newell, working in the research laboratory of Prof. John C Gray at the University of Cambridge, U.K., and with Mr. Peter Cain, under the supervision of Prof. Colin Robinson at the University of Warwick, U.K. These results are described in chapter 4 and concern experiments conducted to determine the sub-cellular location of the Chloroplast Sensor Kinase (CSK). This thesis also incorporates results of experiments undertaken in collaboration with Mr. Lars Dietzel, working under the supervision of Dr. Thomas Pfannschmidt at Friedrich-Schiller-Universität, Jena, Germany. This latter collaboration was directed at characterising the chloroplast transcriptional response of two CSK null mutants and results are presented in chapter 5.

I certify that, with the above qualifications, this thesis, and the research it describes, are entirely the product of my own work, and any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

Sujith Puthiyaveetil
26 June 2008

Abstract

Two-component signal transduction, comprising sensor kinases and response regulators, is the predominant signalling mechanism in prokaryotes. This signalling system originated in bacteria, and has spread to the eukaryotic domain of life through symbiotic, lateral gene transfer from the bacterial ancestors of chloroplasts and mitochondria. During the course of their evolution, chloroplasts, with the exception of a few instances in non-green algae, appear to have relinquished all genes encoding two-component systems to their eukaryotic host cell nuclei. In green algae and plants, chloroplast genes for two-component systems were neither known nor were chloroplast two-component proteins shown to exist as products of nuclear genes prior to the work described here. This thesis describes the identification and characterisation of a novel two-component sensor kinase in chloroplasts. This *C*hloroplast *S*ensor *K*inase (CSK) is the product of a nuclear gene in algae and plants. CSK is synthesised in the cytosol of *Arabidopsis thaliana* and imported into the chloroplast as a protein precursor. CSK is autophosphorylated and couples photosynthetic electron transport to gene transcription in chloroplasts. The identity of the response regulator partner of CSK reveals an unexpected phylogenetic and functional relatedness of CSK with chloroplast two-component systems of non-green algae. Chloroplast two-component systems are likely to be universal in photosynthetic eukaryotes and they persist in chloroplasts as products of nuclear genes even where chloroplast genomes no longer encode them. Chloroplast two-component systems have homologues in extant cyanobacterial lineages, indicating their ancient cyanobacterial origin. The persistence of cyanobacterial two-component systems in chloroplasts and their function in coupling photosynthesis with chloroplast gene expression are central to the premise that chloroplasts retain genes whose expression is regulated by the activity of the photosynthetic electron transport chain, using a mechanism conserved from their cyanobacterial ancestors.

This thesis is dedicated to my parents

Acknowledgements

I thank my supervisor John F Allen for not filtering my email request for a PhD studentship as a "graduate spam," for his inspiring hypothesis that set the stage of my PhD research, for giving freedom to define and perform my own research and for sharing his view that jumping beyond the evidence is essential for discoveries in science. I thank Carol Allen for her support and discussions; Jörgen Ström for helpful guidance in the lab when I first started in Lund; Prof. Conrad Mullineaux and his lab members for help and assistance with techniques in molecular biology and for sharing resources.

I thank Prof. Tony A Kavanagh, Dr. Christine Newell and Prof. John C Gray (University of Cambridge), Mr. Peter Cain and Prof. Colin Robinson (University of Warwick) for a highly productive and decisive collaboration that clinched the sub-cellular location of CSK. In particular, I thank Tony A Kavanagh and Peter Cain for supplying high-quality images of the localization results to include in this thesis. My thanks also go to Mr. Lars Dietzel and Dr. Thomas Pfannschmidt for hosting my stay in Jena and for collaborative experiments that strengthened the conclusion of this thesis. I especially thank Lars for his hospitality, careful experiments and for supplying data that are incorporated in this thesis.

I am thankful to Matthew Rogers (Exeter University) for help with phylogenetic trees; Mark van der Giezen (Exeter University) for help and advice on cloning and for support and encouragement; Yumiko Tashiro-Thatcher, Arefeh Seyedarabi, Thomas Hutchison and Prof. Richard Pickersgill, for advice and assistance with overexpression of CSK; Dr. Brendan Curran for being always cheerful, supportive and for providing useful tips on research and thesis writing; Dr. K. Raveendran (Sir Syed college, Kannur, Kerala) for encouraging me to take up a career in research and for his passionate advocacy of the theory of evolution.

I thank my brothers Prasoon and Sarath, for their support and encouragement. I am grateful to my wife Shanitha for her love, support, and patience during these three and half years of my PhD research. One of the best experiences that we lived through in this period was the birth of our daughter Malavika, who provided an additional, joyful dimension to our life. I acknowledge Queen Mary, University of London for a postgraduate research studentship.

Abbreviations

ABRC	arabidopsis biological resource center
ADP	adenosine 5'-diphosphate
Arc	anoxic redox control
ATP	adenosine 5'-triphosphate
ATPase	ATP synthase
a.u.	arbitrary units
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumin
CaMV	cauliflower mosaic virus
cDNA	complementary DNA
CES	control by epistasy of synthesis
Chl	chlorophyll
CORR	*co*-location for *r*edox *r*egulation
CSK	chloroplast sensor kinase
C_T	cycle threshold
Cyt b_6f	cytochrome b_6f complex
Cyt b_6	cytochrome b_6
Cyt *f*	cytochrome *f*
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid
EST	expressed sequence tag
F_A	iron-sulphur centre F_A
F_B	iron-sulphur centre F_B
Fd	ferredoxin
FNR	ferredoxin $NADP^+$ oxidoreductase
Fx	iron-sulphur centre Fx
GFP	green fluorescent protein
GST	glutathione s-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hik	histidine kinase
Hpt	his-containing phosphotransfer domain
IPTG	isopropyl β-D-1-thiogalactopyranoside
JGI	joint genome institute
kbp	kilobase pair
kDa	kiloDalton
LHC I	light harvesting complex I
LHC II	light harvesting complex II
M	molar
μE	microEinsteins

μg	microgram
μl	microlitre
μm	micrometre
μM	micromolar
ml	millilitre
min	minute
mRNA	messenger RNA
Mt	CSK null mutant
$NADP^+$	nicotinamide adenine dinucleotide phosphate oxidised form
NCBI	national center for biotechnology information
ng	nanogram
OD	optical density
PC	plastocyanin
PCR	polymerase chain reaction
PEP	plastid encoded polymerase
PQ	plastoquinone
PQH_2	plastoquinol
PS I	photosystem I
PS II	photosystem II
PSI BLAST	position specific iterative blast
qPCR	quantitative PCR
RCF	relative centrifugal force
redox	reduction-oxidation
RNA	ribonucleic acid
RNase	ribonuclease
rpm	rotation per minute
rre	response regulator
RT-PCR	reverse transcriptase PCR
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
S	seconds
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	standard error
SMART	simple modular architecture research tool
TAIR	the arabidopsis information resources
TCP34	tetratricopeptide containing chloroplast protein of 34 kDa
Tris	tris(hydroxymethyl)aminomethane
UTP	uridine 5'-triphosphate
UV	ultraviolet light
Wt	wild type
ycf	hypothetical chloroplast open reading frame

Table of contents

List of figures

List of tables

1.1 Two-component signal transduction: from obscurity to prominence

Underscoring biochemical unity in living organisms, the pre-eminent French biologist Jacques Monod once famously quipped: "What's true for *E. coli* is true for an elephant." Monod's aphorism holds in the case of phosphorylation. Reversible phosphorylation is a fundamental regulatory mechanism present in all domains of life (Anderson, 1992, Pawson and Scott, 2005). Living systems must have therefore learned the trick of sticking on a phosphate group for regulation very early in evolution, possibly extending deep in time at the hydrothermal hatcheries of life. Acetyl phosphate, probably the first organic phosphate formed in the chemoautotrophic metabolism of the protocell (Russell and Hall, 2006), could have been the earliest phosphoryl group donor (Duve, 1991). Acetate kinase, which catalyses the reversible conversion of acetyl phosphate and ADP into acetate and ATP, could be the earliest phosphotransferase (Buss et al., 2001), the first of the many enzymes that transfer phosphate groups. Although acetyl phosphate continues to function as an important phosphoryl group donor, this role has now largely been taken over by the energy rich compound Adenosine Triphosphate (ATP). Multitudes of phosphotransferases, which include protein kinases, have also evolved since.

Protein kinases are enzymes that catalyze the transfer of the γ-phosphate from ATP or GTP onto an amino acid side chain of a protein substrate. At least five major amino acid phosphorylation sites have been recognized in proteins. Phosphorylation on serine is the most common, followed by threonine. Tyrosine phosphorylation is relatively rare, while histidine phosphorylation is estimated to be 10 to 100 fold more abundant than phosphotyrosine (Klumpp and Krieglstein, 2002). It was once believed that serine/threonine/tyrosine phosphorylation was a defining feature of eukaryotes and histidine/aspartate phosphorylation of prokaryotes. Recent research, including that presented in this thesis, shows that this dichotomy no longer exists.

Histidine and aspartate phosphorylation occur as parts of a signalling pathway known as two-component signal transduction (Stock et al., 2000). Two-component

signalling systems have come to be known only relatively recently (Nixon et al., 1986, Stock et al., 1985). This is partly because signal transduction research has thus far focussed heavily on serine/threonine/tyrosine kinases, the so-called "eukaryotic kinases" (Manning et al., 2002). Histidine phosphorylation has also been overlooked because of the lability of histidine phosphates in traditional biochemical assays (Klumpp and Krieglstein, 2002). Two-component signalling is now an active area of research. Every now and then a new two-component system comes to light. The name "two-component system" reflects its experimental history as this regulatory system was first identified as two conserved components in genetic screens of bacteria (Robinson et al., 2000). Two-component systems are widespread in bacteria, though not ubiquitous. Bacteria such as *Escherichia coli* have around 32 two-component systems (Mizuno, 1997), while some parasitic bacteria like *Mycoplasma genitalium* have none (Mizuno, 1998). Some cyanobacteria such as *Anabaena* sp. strain PCC 7120 contain around 211 two-component system genes (Wang et al., 2002).

1.2 Two components make a signal transduction chain

Like most signalling pathways, two-component signal transduction is built around conserved components. The conserved components that form the core of the two-component signalling pathway are a sensor kinase and a response regulator (Figure 1.1) (Stock et al., 2000). In a two-component signal transduction chain, the sensor kinase lies at the interface between the cellular environment and the adaptive response (Figure 1.1). The sensor kinase is a histidine protein kinase, which in its functional dimeric form undergoes an ATP dependent trans-autophosphorylation reaction whereby one histidine kinase monomer phosphorylates a second monomer within the dimer, on a conserved histidine residue. The phosphoryl group from the histidine kinase is then transferred to a conserved aspartate residue on the response regulator protein (Figure 1.1). Sensor histidine kinases thus differ from serine/threonine kinases in that the latter kinase family transfers the phosphate group from ATP directly onto the substrate.

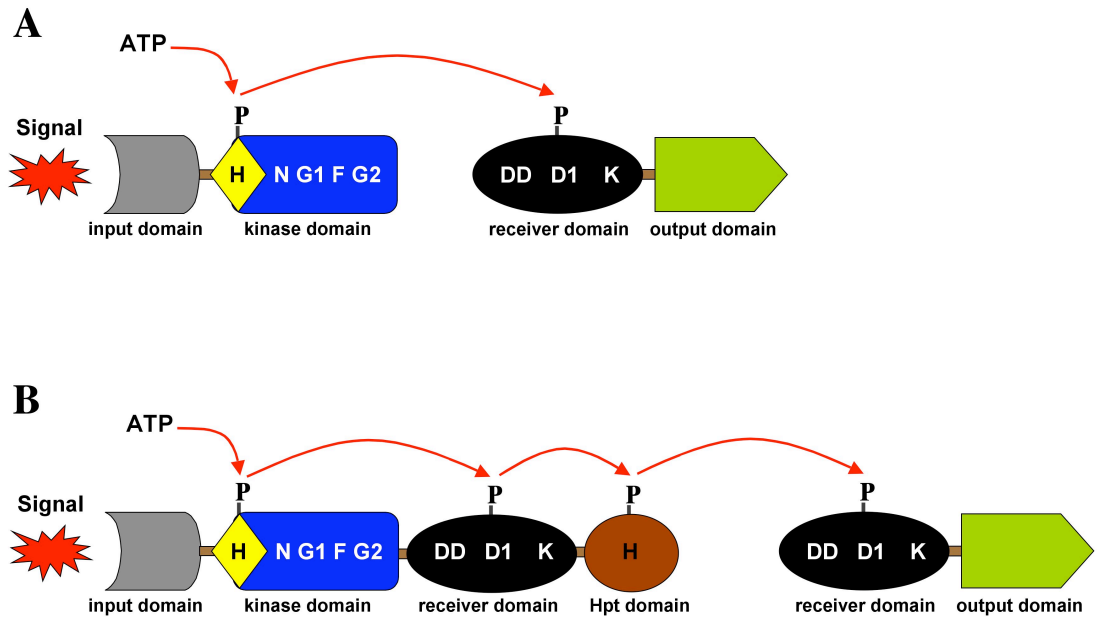


Figure 1.1 Two-component signalling systems: their modular design and modes of signalling. (**A**) A phosphotransfer pathway involving an orthodox sensor kinase and a response regulator that participate in a one-step His-Asp phosphoryl group transfer. The kinase domain of the sensor consists of a dimerization domain (yellow diamond) and an ATP binding domain (rounded blue rectangle). The H, N, G1, F and G2 motifs of the kinase domain are indicated. Conserved sequence motifs of the response regulator receiver domain, DD, D1 and K are also indicated. (**B**) A phosphorelay pathway involving a hybrid kinase and a response regulator that together participate in a three-step phosphoryl group transfer. The Hpt domain is the His-containing phosphotransfer domain.

Sensor histidine kinases combine a variable sensor domain with an invariable kinase domain (Figure 1.1) (Stock et al., 2000). Most sensor histidine kinases are transmembrane proteins with the sensor domain located in the extra-cytoplasmic space and the kinase domain in the cytoplasm. The N-terminal, sensor domain of the histidine kinase perceives the signal or stimulus that controls the activity of the kinase domain. The nature of the signal sensed and the protein's sensor domain are unique for each histidine sensor kinase. In contrast, the kinase domain is highly conserved in structure and function, being made up of an independent dimerization motif and a catalytic core (Figure 1.1). The catalytic core of the kinase domain consists of five conserved amino acid motifs: H-box, N, G1, F and G2 (Figure 1.1 and figure 1.2). The H-box contains the conserved histidine residue that is the site of phosphorylation, and is usually located in the dimerization motif. N, G1, F and G2 boxes form the ATP binding pocket of the catalytic core (Figure 1.1) (Stock et al., 2000).

Structural information is available for the kinase domain of a few histidine kinases (Bilwes et al., 1999, Tanaka et al., 1998). From the structure, it was found that the kinase domain of histidine kinases is unlike that of any previously known kinase fold (Figure 1.2). However, similarity was noted in the ATP binding domain of histidine kinase with a family of ATP hydrolases, which includes DNA gyrase B, heat shock protein Hsp90 and the DNA-mismatch-repair protein MutL. The ATP binding fold characteristic of these hydrolases and histidine kinases is called a Bergerat fold (Figure 1.2) (Bergerat et al., 1997). It is believed that a slight variation in the "ATP lid" of a Bergerat fold determines whether the protein exhibits a histidine kinase activity or hydrolase activity (Dutta and Inouye, 2000). Each dimerization or His-containing domain within the kinase domain is made up of two antiparallel helices that form a coiled coil structure (Figure 1.2) (Stock et al., 2000). A dimeric histidine kinase is formed through the association of these coiled coil structures from two histidine kinase monomers, thus creating a four-helix bundle (Figure 1.2). The conserved histidine residue that is the site of autophosphorylation

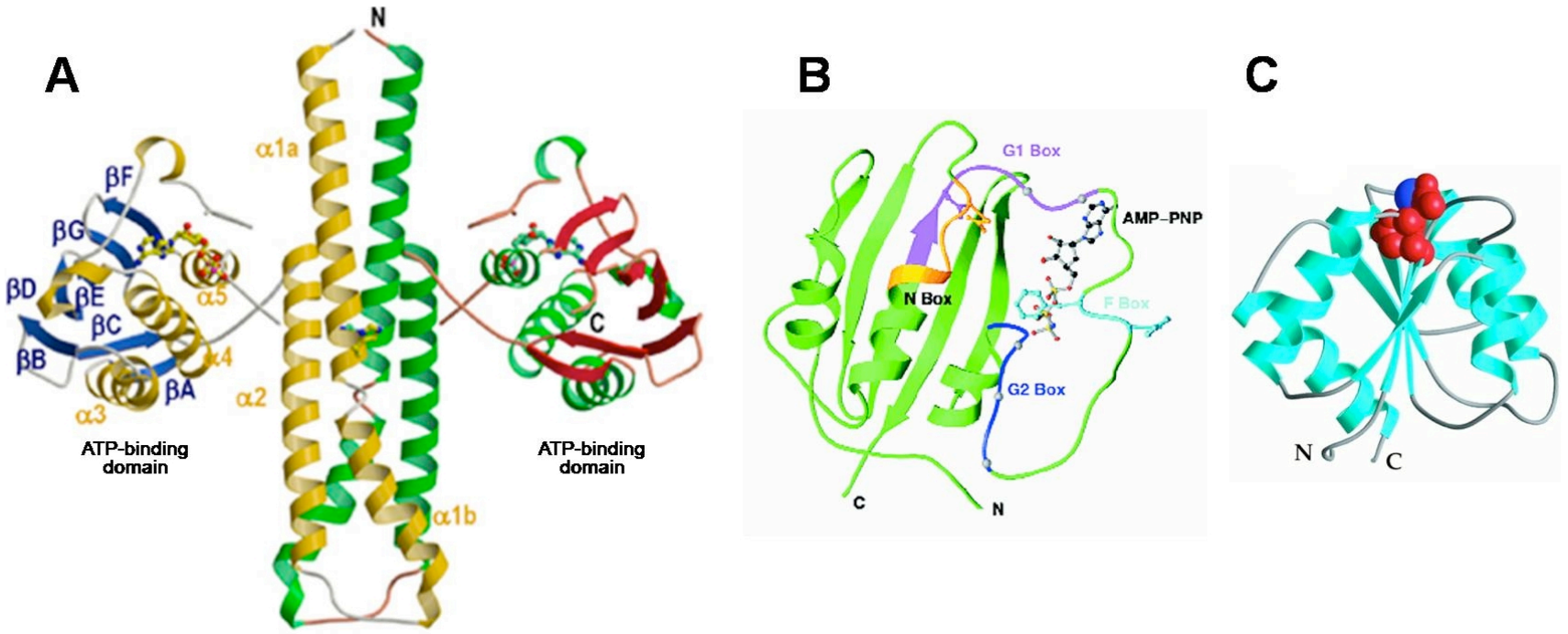


Figure 1.2 Conserved domains in sensor histidine kinases and response regulators. (**A**) The Kinase domain of *Thermotoga maritima* sensor histidine kinase TM0853 in dimeric form. *Alpha* helices are labelled α1-α5, and coloured gold for monomer A and green for monomer B. *Beta* sheets are labelled βA-βF and coloured blue for monomer A and red for monomer B. *Alpha* helices, α1 *and* α2 from each monomer form the dimerization domain. The ATP binding domain (labelled) is an α/β fold. (**B**) ATP binding domain of *E. coli* osmosensor EnvZ with a non-hydrolysable ATP analog AMP-PNP. Conserved amino acid motifs, N, G1, F and G2, that form the ATP binding pocket are highlighted. (**C**) Receiver domain of *Salmonella typhimurium* chemotaxis protein CheY. The α/β fold encloses a central five-stranded parallel β sheet that is surrounded by five α helices. The conserved cluster of aspartic acid residues (red spheres) bind a Mg^{2+} (dark blue) and form the active site for phosphoryl group transfer. (Figure adapted from (Robinson et al., 2000, Stock et al., 2000, Marina et al., 2005)).

is positioned midway on the solvent-exposed side of the two opposing helices in the four-helix bundle (Stock et al., 2000).

Phosphorylation on nitrogen of the histidine residue in histidine kinases produces a phosphoramidate linkage that is different from phosphoester linkages of the serine/threonine/tyrosine family of kinases. The phosphoramidate bond has a large negative free energy of hydrolysis when compared to phosphoester linkages and this makes phosphohistidine the most unstable of any known phosphoaminoacid (Klumpp and Krieglstein, 2002). At typical intracellular ATP/ADP ratios, the equilibrium of histidine kinase autophosphorylation favours the unphosphorylated form. It is assumed that histidine kinases in their resting phase have a constitutively active catalytic core that is out of contact with the substrate histidine residue. But, when an appropriate signal is detected by the sensor domain, conformational changes bring the catalytic core of one monomer close to the conserved histidine residue of the other monomer, allowing phosphorylation of the substrate histidine residue (Saito, 2001). Most histidine kinases, in addition to acting as kinases, also function as response regulator phosphatases (Robinson et al., 2000). This bifunctional action regulates the phosphorylation state of downstream response regulators, and so can be used to control the flow of information through the signalling pathway.

Some members of the histidine kinase family are known as hybrid kinases and these have additional functional domains that act as phosphoacceptor and phosphodonor sites (Figure 1.1) (Stock et al., 2000). These domains allow additional inputs and control points into the signalling pathway. In hybrid kinases, the phosphoacceptor domain is formed by a receiver domain, which is usually found in response regulators (Figure 1.1). This receiver domain in hybrid kinases receives a phosphoryl group from the phosphohistidine of the kinase domain (Figure 1.1). The phosphodonor domain, on the other hand, is formed by a histidine–containing phosphotranfer domain (Hpt) (Figure 1.1). The Hpt domain receives its phosphoryl group from the phosphoaspartate of the receiver domain in the hybrid kinase. The Hpt domain then donates this phosphate group to the receiver domain of the response

regulator (Figure 1.1). Hpt domains share a four-helix bundle structure that is homologous to the four-helix bundle formed by dimerization domains of two histidine kinase monomers (Stock et al., 2000). As in the dimerization of a histidine kinase, the conserved active site histidine residue in the Hpt domain is also positioned on the solvent- exposed helical face. In prokaryotes, the Hpt domain is usually part of the hybrid kinase (Figure 1.1), while in eukaryotes it is typically found as a separate protein. Hybrid kinases are rare in prokaryotes, but in eukaryotes histidine kinases are mostly in hybrid form (Robinson et al., 2000).

Histidine kinases contain a variety of extracellular, intracellular and transmembrane sensor domains designed for detecting specific ligands or stimuli (Robinson et al., 2000). Sensor domains are highly heterogeneous in size and structure and therefore have little sequence similarity in different histidine kinases. It is believed that when sensor kinases duplicate and diverge, sensor domains acquire new signal recognition functions by a hyper variable sequence evolution, while the kinase domain remains unchanged (Hoch, 2000). Two commonly found cytoplasmic sensor domains in histidine kinases are PAS and GAF. PAS domains, named after the PER, ARNT and SIM proteins, sense signals such as light, redox reactions, and oxygen, usually with the help of an associated cofactor (Taylor and Zhulin, 1999). GAF domains, named after vertebrate cGMP specific phosphodiesterase, cyanobacterial Adenylate cyclase and the bacterial formate hydrogen lyase transcription activator FhlA, are structurally related to PAS domains (Ho et al., 2000). GAF domains bind a variety of small molecule ligands, including cyclic nucleotides (Zoraghi et al., 2004), bilins (Montgomery and Lagarias, 2002), 2-oxoglutarate (Little and Dixon, 2003), sodium ions (Cann, 2007) and nitric oxide (D'Autreaux et al., 2005). GAF domains also sense redox signals via redox-responsive prosthetic groups such as haem (Kumar et al., 2007).

The second component of any two-component system is its response regulator (Figure 1.1). Most response regulators have a two-domain architecture (Stock et al., 2000). The N-terminal domain is the invariable receiver domain that

catalyses the transfer of a phosphate group from the phosphohistidine of the sensor kinase to one of its own aspartate residue (Figure 1.1). The C-terminal domain in response regulators is the variable effector domain that mediates the specific output response (Figure 1.1). The receiver domain contains three conserved sequence motifs: DD, D1 and K. The D1 motif is the aspartate residue that is the site of phosphorylation (Figure 1.2). The DD motif is a pair of aspartate residues involved in metal binding (Figure 1.2) and the K motif is a lysine residue that interacts with the phosphate (Robinson et al., 2000). The histidine kinase–mediated phosphorylation on the conserved aspartate residues in the receiver domain of response regulators creates a high-energy acyl phosphate. This phosphorylation event causes the activation of the effector domain by a conformational change that is propagated from the receiver domain to the effector domain (Robinson et al., 2000). The exact details of this conformational change remain to be resolved.

Response regulators bring about a diverse set of output responses depending on the nature of their effector domains. The most common output response is the regulation of gene expression, owing to the transcriptional regulatory property of effector domains. Based on homology in their effector domains, three subfamilies of transcriptional regulatory response regulators are recognized (Robinson et al., 2000). These subfamilies are OmpR, NarL and NtrC, named after representative family members. All subfamilies contain a helix-turn-helix DNA binding motif in their effector domains, but the overall fold of the effector domain varies between subfamily members. The effector domain in OmpR is a winged helix-turn-helix. NarL has a four–helix bundle structure and NtrC has an ATPase domain in addition to the helix-turn-helix motif (Robinson et al., 2000).

As their signalling mechanism, two-component systems use a phosphoryl group transfer pathway from the invariant kinase domain of the sensor to the invariant receiver domain of the response regulator (Figure 1.1) (Robinson et al., 2000). Two basic modes of phosphoryl group transfer pathway exist in two-component signal transduction. The simplest one is the phosphotransfer pathway involving a single

histidine kinase and a single response regulator that together participate in a one step His→Asp phosphoryl group transfer (Figure 1.1). A more complex form is the phosphorelay pathway. This pathway usually involves a hybrid kinase and a response regulator protein that participate in a three step His→Asp→ His→Asp phosphoryl group transfer (Figure 1.1). In the phosphorelay pathway, the Hpt domain, which function as a phosphodonor domain, is seen either as part of the hybrid kinase or as a separate protein. Simple phosphotransfer pathways predominate in prokaryotes, while phosphorelays, which integrate multiple steps of regulation, are frequent in eukaryotes.

1.3 Two-component systems are not confined to prokaryotes

Up until a decade ago, two-component systems were though to be unique to prokaryotes. However, a flurry of activity in both genomics and molecular genetics has dispelled this notion of uniqueness. Beginning with the discovery of ethylene receptors in plants (Chang et al., 1993), two-component systems have since been reported in various eukaryotic organisms, including fungi, slime moulds, unicellular algae, mosses and higher plants (Saito, 2001, Thomason and Kay, 2000). It is believed that two-component systems originated in bacteria but have spread to other domains of life through lateral gene-transfer (Koretke et al., 2000). A simple lateral gene transfer event has taken two-component system to archaea, but eukaryotes received two-component systems from the bacterial ancestors and early evolutionary precursors of chloroplasts and mitochondria through symbiotic gene transfer, a special case of lateral gene transfer. Among eukaryotes, plant, moss, and algal genomes encode many two-component systems (Hwang et al., 2002, Rensing et al., 2008), while in fungi, only a few two-component systems are found (Catlett et al., 2003).

Two-component systems are considered to be absent in animals. Their absence has been variously proclaimed. As one title (Wolanin et al., 2002) reads “Histidine protein kinases: key signal transducers outside the animal kingdom.” For this reason,

histidine kinases have been perceived as a potential drug-target against bacteria and fungi (Stephenson and Hoch, 2002). Recent research, however, suggests that it may be premature to dismiss the presence of two-component systems in animals. Studies have now looked at two well-known mitochondrial enzymes, branched-chain α-ketoacid dehydrogenase kinase (BCKDHK) (Lasker et al., 2002) and pyruvate dehydrogenase kinase (PDHK) (Lasker et al., 2002, Mooney et al., 2000, Thelen et al., 2000). These enzymes have high sequence similarity to sensor histidine kinases, but have nevertheless been classified as serine/threonine kinases. From the above studies, it emerged that both these enzymes, in addition to their serine/threonine kinase activity, also possess a histidine kinase activity. This was rather a surprising revelation. Even though histidine phosphoproteins have been reported in many animal studies, this is the first demonstration that proteins that look like histidine kinases also behave like histidine kinases in animals. Sequence similarity search shows response regulator homologues in some metazoans such as *Mus musculus* and *Nematostella vectensis* (unpublished research, Puthiyaveetil and Allen). This may prove interesting, but sequence contamination cannot be ruled out at this stage. As discussed in chapter 3, the chemical instability of phosphohistidines in biochemical assays and sequence modification of sensor histidine kinases and response regulators can impede the identification of two-component systems in metazoans.

1.4 Functional roles of two-component systems: from motility to virulence

Two-component systems regulate diverse adaptive responses in bacteria, fungi and plants (Bekker et al., 2006, Santos and Shiozaki, 2001, Hwang et al., 2002). Chemotaxis in bacteria is a well-characterised two-component sensory mechanism. Other functions of two-component systems in bacteria include acquisition of nutrients such as nitrogen, phosphate and carbon, and micro and macro elements like nickel and magnesium; regulation of respiration and photosynthesis; sensing of chemical and physiological stimuli such as pH, osmolarity, low temperature and light; controlling developmental processes such as sporulation; cell-to-cell communication, quorum sensing; and control of virulence. In fungi, two component

systems regulate processes such as osmosensing and oxidative stress. In slime moulds, fruiting body formation is under the control of two-component systems (Wang et al., 1996). In plants, two-component systems function as osmosensors, hormone receptors and light receptors.

A number of responses controlled by two-component systems in bacteria involve redox-regulated gene transcription. In such cases, sensor kinases function as redox sensors and are mostly wired into photosynthetic or respiratory electron transport chain (Allen, 1993b, Bauer et al., 1999, Green and Paget, 2004). A well-known example is the ArcAB system, which functions as a regulatory switch that selects between aerobic and anaerobic metabolisms in bacteria. Redox state of the quinone electron carrier is the regulatory signal for the ArcB sensor (Georgellis et al., 2001). The RegAB/PrrAB system in photosynthetic bacteria is also a quinone-level redox switch that acts as a global regulator of metabolism, including photosynthesis, respiration and nitrogen fixation, depending on the availability of oxygen (Swem et al., 2006). The NifAL system in the nitrogen fixing bacterium, *Klebsiella pneumoniae* controls the expression of *nif* genes according to the partial pressure of oxygen. The NifL sensor monitors oxygen tension indirectly from the redox state of the quinone pool (Grabbe and Schmitz, 2003). The BvgS and EvgS sensor kinases of *Bordetella pertussis* and *Escherichia coli*, respectively, are also known to sense quinone redox state (Bock and Gross, 2002). Quinone sensing by BvgS and EvgS is likely to modulate virulence in these bacteria. The DosS sensor of *Mycobacterium tuberculosis* is also a redox sensor that tunes the expression of virulence genes (Kumar et al., 2007).

1.5 Chloroplasts

Chloroplasts are cytoplasmic organelles that originated from endosymbiotic cyanobacteria (Raven and Allen, 2003). They are the sites of photosynthesis in plants and eukaryotic algae. These disc-shaped organelles are double-membrane bound and contain an intricate membranous system of flattened vesicles, called thylakoids,

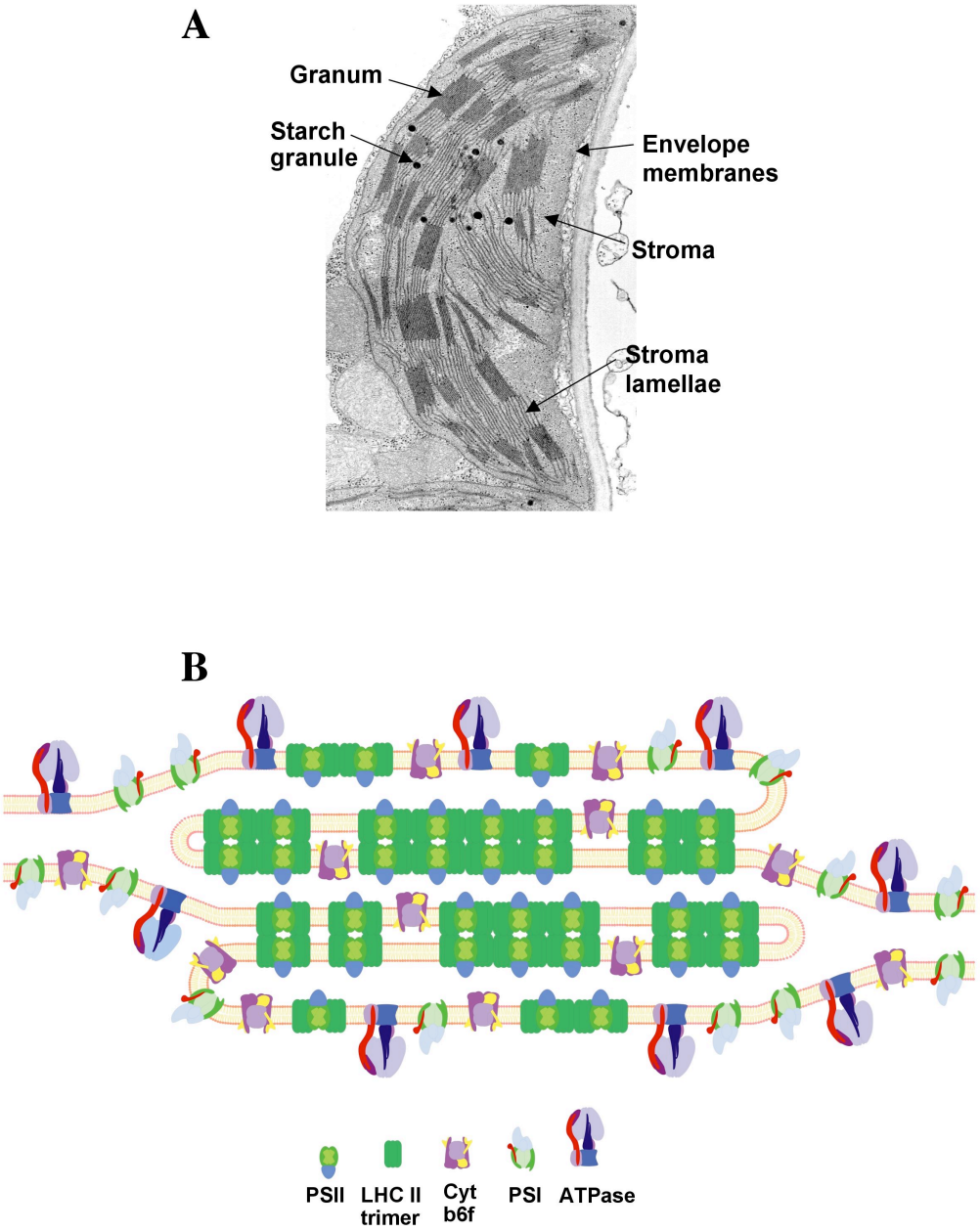


Figure 1.3 Ultrastructure of chloroplasts and organization of electron transfer complexes. (**A**) Electron micrograph of a chloroplast showing the organization of thylakoids into grana and stroma lamellae. (**B**) Diagram depicting the spatial organization of electron transfer complexes in the thylakoid membrane. (Figure 1.3A taken from image collections of the Department of Botany, UW-Madison. Figure 1.3B based on (Allen and Forsberg, 2001))

suspended in a jelly-like fluid known as stroma (Figure 1.3A) (Kirk and Tilney-Bassett, 1978). Thylakoids are organized into stacks of disks forming grana, which are interconnected by single, unstacked membranous structures known as stroma lamellae (Figure 1.3A). A chloroplast contains special light absorbing pigment molecules called chlorophylls and carotenoids, studded onto proteins and arranged in special light harvesting and energy converting centres called photosystems (Figure 1.3B). Photosystems and other complexes necessary for the light harvesting and electron transfer functions of photosynthesis are usually plugged into the thylakoid membrane (Figure 1.3B). Photosystems contain a ring of light harvesting peripheral antenna molecules and a reaction centre core containing a special pair of chlorophyll molecules. There are two such photosystems in a chloroplast, one preferentially absorbing light of shorter wavelength (up to 680 nm) called photosystem II (PS II) and the other absorbing longer wavelength of visible spectrum (up to or beyond 700 nm) known as photosystem I (PS I) (Figure 1.3B). These two different wavelength absorbing photosystems are joined together for electron transfer by a complex called cytochrome b_6f (Cyt b_6f) (Figure 1.3B).

Chloroplasts convert light energy from the sun into chemical energy, in the form of carbohydrates. Apart from light, the essential raw materials for this reaction of global significance, and on which, ultimately, all life forms depend for energy, are CO_2 and H_2O. Water is used as an electron source. The electrons stripped from water are the reducing power, required for producing carbohydrates from CO_2. The proton gradient generated in the process is used to produce ATP. O_2 is a by-product of the water splitting reaction in photosynthesis. The advent of oxygenic photosynthesis was an unparalleled evolutionary innovation in the history of life (Allen and Martin, 2007). Oxygenic photosynthesis generated an oxygenic atmosphere and allowed aerobic respiration, which fuelled the origin and evolution of multicellular organisms (Lane, 2002).

The process of photosynthesis is divided into two main sets of reactions: light (light-dependent) reactions and dark (light-independent) reactions. Light reactions

are a series of light-driven electron transfer reactions performed by photosystems (Figure 1.4). Most chlorophyll molecules present in photosystems function in light–harvesting antenna that increase the surface area over which light energy is absorbed. These chlorophylls and accessory pigments in the light–harvesting complex absorb each quantum of light, and the resulting excitation energy is transferred into the reaction centres of each photosystems. Reaction centres use the excitation energy received from light–harvesting pigments to drive a charge separation reaction between a special chlorophyll species and an electron acceptor. The chlorophyll species is therefore a primary electron donor, and is called P680 in photosystem II (PS II) and P700 in photosystem (PS I), (Figure 1.4). These special chlorophyll molecules thus loose the electrons to the acceptor and are thus oxidized.

Electrons generated during the charge separation in PS II are passed along a series of electron acceptors to reduce the plastoquinone (PQ) to plastoquinol (PQH_2) (Figure 1.4). PQH_2 is re-oxidised by the cytochrome (cyt) b_6f complex. Re-oxidation of each PQH_2 is coupled with the movement of four protons from the stroma into the lumen of the thylakoid. This proton translocation reaction by cyt b_6f complex contributes in generating a proton gradient across the thylakoid membrane, which is used by the ATPase enzyme to make ATP. Electrons derived from the oxidation of PQH_2 by cyt b_6f are passed onto a soluble electron acceptor called plastocyanin (PC) (Figure 1.4). PC diffuses into PS I and donates the electrons to PS I to reduce the oxidised P700. Electrons generated from the light–induced charge separation of P700 reduce ferredoxin (Figure 1.4). Electrons from ferredoxin reduce $NADP^+$ to NADPH by ferredoxin–$NADP^+$ reductase (FNR) (Figure 1.4). This linear electron transport chain is completed by the reduction of the P680 special-pair in PS II via the oxidation of H_2O.

In the electron transport reactions of photosynthesis, each electron carrier undergoes a cycle of chemical reduction and oxidation, becoming reduced by accepting an electron from the previous carrier, and then oxidised by donating the

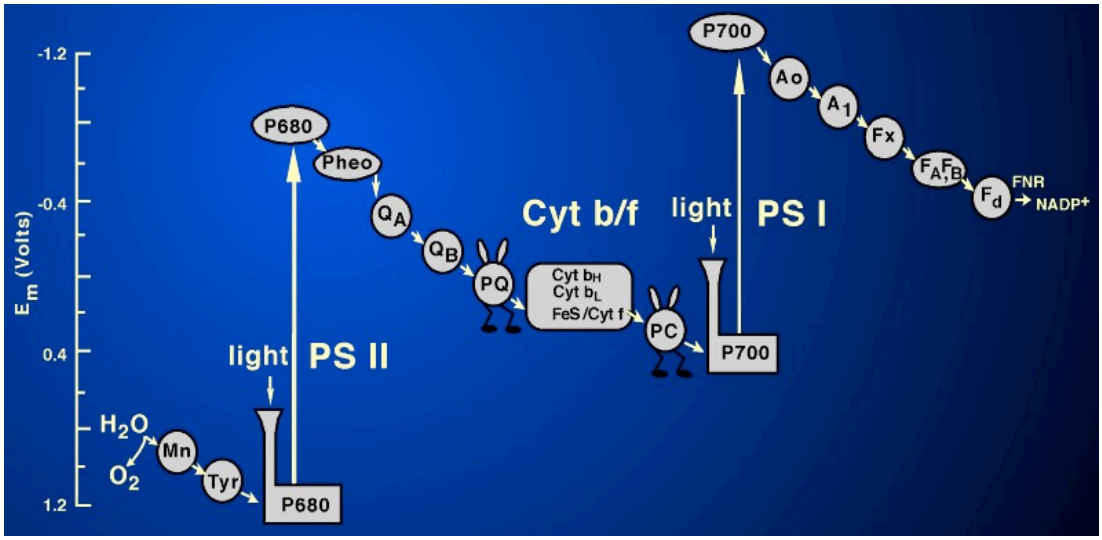


Figure 1.4 The Z-scheme of photosynthesis. The electron transport pathway from H_2O to $NADP^+$ is plotted on a scale of redox potential. "E_m" is the midpoint potential expressed in volts. Abbreviations: Mn, manganese cluster; Tyr, tyrosine; P680, PS II reaction centre special pair chlorophyll; Pheo, Phaeophytin; Q_A, plastoquinone A (bound plastoquinone); Q_B, plastoquinone B (mobile plastoquinone); PQ, plastoquinone pool; Cyt b/f, cytochrome b_6f complex; PC, plastocyanin; P700, PS I reaction centre special pair chlorophyll; A_0, primary acceptor chlorophyll; A_1, secondary acceptor phylloquinone; Fx, iron sulphur cluster; F_AF_B, iron sulphur cluster; Fd, ferredoxin; FNR, ferredoxin $NADP^+$ oxidoreductase; $NADP^+$, nicotinamide adenine dinucleotide phosphate oxidised form. (Z-scheme taken from (Govindjee, 2000))

electron to the next carrier in the chain (Figure 1.4). This reduction–oxidation or "redox" chemistry underpins energy conversion in photosynthesis. In the electron transport chain, each electron carrier is positioned according to their redox potential, which is the affinity for electrons (Figure 1.4). Electrons move spontaneously from a less positive potential to a more positive potential in the electron transport chain (Figure 1.4). However in photosynthesis, for this down-hill flow of electrons to occur, electrons should first be driven up-hill from reaction centres by using the energy from light. If reaction centres and electron carriers are arranged according to their redox potential, they form the Z-scheme of light reactions (Figure 1.4) (Hill and Bendall, 1960). The scheme derives its name from the resemblance to the letter "Z" when it was first drawn.

In the dark reactions of photosynthesis, free energy from the hydrolysis of ~P bonds of ATP and the reducing power of NADPH, both formed during the light reactions, are used to reduce CO_2 to sugars through a series of biochemical reactions called the Calvin cycle, the Benson-Calvin cycle, or the reductive pentose phosphate pathway. Enzymes and the intermediates of Calvin cycle are located in the stroma. The key enzyme in the Calvin cycle is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which is also one of the most abundant proteins on Earth. The term "Dark reaction" is something of a misnomer, these reactions neither require darkness nor do they operate in darkness. They are, themselves, merely light-independent. Dark reactions, on the contrary, also require light indirectly for the activation of many of the enzymes participating in the Calvin cycle (Buchanan, 1980).

Four major multisubunit complexes, PS I, PS II, cyt b_6f and ATPase (Figure 1.3B), operate in the thylakoid membrane for capturing and converting light energy into the reductants and ATP required for fixing CO_2 to carbohydrates. In addition to these four major complexes, thylakoid membranes also contain remnants of the respiratory electron transport chain of the ancestral cyanobacterium. A plastid-encoded NAD(P)H dehydrogenase (ndh) complex, homologous to the bacterial

complex I, and a recently identified nuclear-encoded plastid terminal oxidase (PTOX), homologous to the plant mitochondrial alternate oxidase, are thought to be part of a chlororespiratory pathway in chloroplasts (Peltier and Cournac, 2002).

PS II is a homodimeric multisubunit pigment-protein-cofactor complex embedded in the thylkoid membrane (Figure 1.5A). PS II is located primarily in the granal region of the thylakoid membrane, where it functions as a water-plastoquinone oxidoreductase. Each PS II monomer consists of around 19 subunits, 16 of which are membrane intrinsic (Figure 1.5A) (Ferreira et al., 2004). Protein subunits of PS II can be separated into four main structural units: a reaction centre core, an inner core of light harvesting antenna, a peripheral ring of light harvesting antenna and a hydrophilic, lumenal oxygen evolving complex. The reaction centre core that lies at the heart of PS II is formed by two homologous proteins known as D1 and D2, which have five transmembrane helices each. In photosynthetic eukaryotes, both D1 and D2 subunits are universally encoded in the chloroplast genome (*psbA* and *psbD* respectively). D1 and D2 subunits bind pigments and cofactors involved in charge separation and electron transfer. CP47 and CP43 subunits (products of chloroplast genes, *psbB* and *psbC* respectively) form the core antenna of PS II (Figure 1.5A). CP47 and CP43 subunits, each having 6 transmembrane helices, bind 16 and 14 chlorophylls respectively (Ferreira et al., 2004). These subunits are thought to be involved in transferring excitation energy from peripheral antenna to the reaction centre core. A number of single membrane-spanning subunits, many of them chloroplast-encoded, are also present in the core of the PS II (Figure 1.5A).

The peripheral light-harvesting antenna keeps the reaction centre core supplied with excitation energy. The composition of this peripheral antenna varies among different organisms. In plants and green algae, this antenna system is membrane–intrinsic and made up of Light Harvesting Complex II (LHC II) proteins that bind chlorophyll *a* and chlorophyll *b*. In cyanobacteria and red algae,

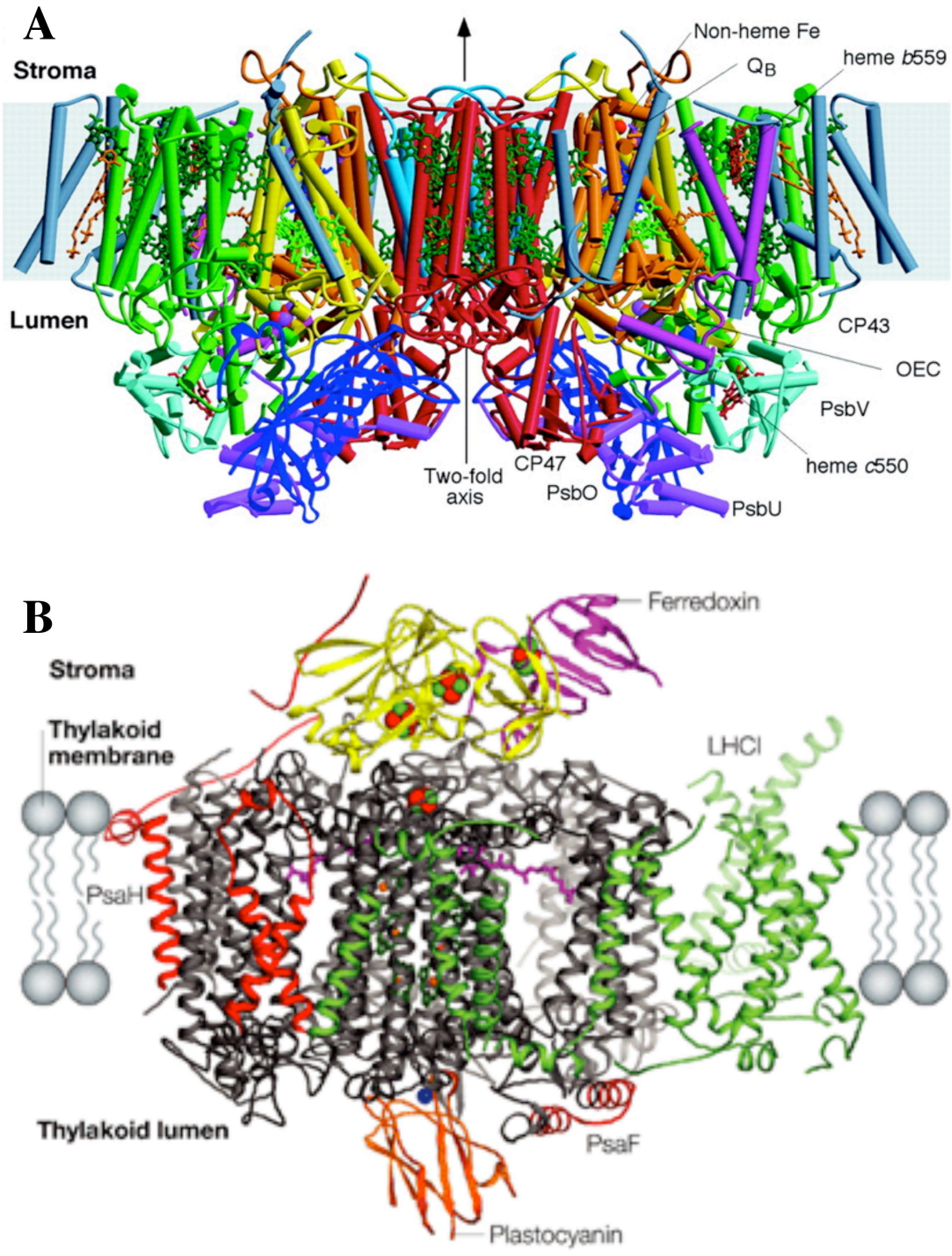


Figure 1.5 Architectures of photosystems. (**A**) The basic structural organization of the dimeric PS II complex from the cyanobacterium *Thermosynechococcus elongatus* viewed parallel to the membrane plane. α-helices are represented as cylinders. (**B**) Structural organization of plant (*Pisum sativum*) PS I

viewed parallel to the membrane plane. α-helices are represented as helical ribbons, β-strands as twisted ribbons and turns and loops as ropes. (Figure adapted from (Ferreira et al., 2004, Nelson and Ben-Shem, 2004))

membrane–extrinsic phycobilisomes, which contain phycobilins and phycoerythrins as light absorbing pigments, form the light harvesting antenna (Glazer, 1985). As many as 300 light–harvesting pigment molecules serve one PS II reaction centre. The oxygen evolving complex, which lies at the lumenal side of PS II, is formed by three extrinsic subunits that house a catalytic Mn_4Ca cluster (Figure 1.5A). The D1 protein provides the majority of the ligands that coordinate the Mn_4Ca cluster. The Mn_4Ca cluster undergoes four stepwise oxidation reactions to supply electrons to the oxidised P680 in the PS II reaction centre. This creates a strongly positive redox potential in the Mn_4Ca cluster, which enables it to split two molecules of water into four electrons, 4 protons and one molecule of dioxygen (Yano et al., 2006).

PS I is a large multisubunit pigment-protein-cofactor complex, which is usually found in the stroma lamellae region of the thylakoid membrane. PS I functions as a plastocyanin-ferredoxin oxidoreductase (Figure 1.5B). PS I generates the most negative redox potential in nature and could be counted as one of the most efficient energy converting machine ever made (Nelson and Yocum, 2006). PS I is a monomeric enzyme in plants and algae, while in cyanobacteria, it functions as a trimer. Each monomeric PS I complex consists of 17 different subunits (Amunts et al., 2007). The core reaction centre of PS I is formed by a heterodimer of PsaA and PsaB subunits, which each have a relative molecular mass of 83 kDa (Figure 1.5B). Both of these subunits are always encoded in the chloroplast genome in plants and algae. These subunits, each containing 11 transmembrane helices, bind many chlorophylls, including the P700 special chlorophyll pair, carotenoids, and other cofactors that assist in electron transfer reactions (Amunts et al., 2007).

The organization of transmembrane helices in the PsaA-PsaB dimer is similar to the arrangement of helices in D1-D2 and CP43-CP47 subunits of PS II (Hankamer et al., 1999). This similarity has led to the suggestion that the two photosystems may share a common evolutionary origin (Nitschke and Rutherford, 1991). The PsaC protein ligates two iron–sulphur centres that receive electrons derived from the oxidation of P700, and is located on the stromal side of PS I (Figure 1.5B). PsaC subunit along with two other peripheral subunits, PsaD and PsaE, form a stromal ridge that docks the soluble electron acceptor ferredoxin (Figure 1.5B) (Amunts et al., 2007). In plant PS I, a small subunit, PsaH (Figure 1.5B) forms the anchor for the mobile LHC II (Lunde et al., 2000), which migrates between two photosystems as part of an acclimatory process known as state transitions (Allen, 2003a). Cyanobacterial PS I lacks a peripheral antenna, while plant and algal PS I have a peripheral antenna complex made up of Light Harvesting Complex I (LHC I). Four LHC I proteins form a half-moon like structure around one side of the PS I complex (Amunts et al., 2007). Chlorophyll *a*/*b* binding proteins in both LHC I and LHC II are encoded by the *Lhc* multigene family (Jansson, 1999). These proteins show high sequence homology suggesting a common evolutionary origin (Green et al., 1991).

The cytochrome b_6f complex forms the electronic connection between PS II and PS I and functions as a plastoquinol-plastocyanin oxidoreductase (Figure 1.6A). It is also the main contributor to the proton gradient required for the synthesis of ATP. The complex occurs in dimeric form and encloses a central cavity for exchanging plastoquinone (Kurisu et al., 2003, Stroebel et al., 2003). Each monomer consists of four large subunits: cytochrome b_6 (Cyt b_6), subunit IV (Sub IV), cytochrome f (Cyt f) and the Rieske iron-sulphur protein (ISP) (Figure 1.6A). Each monomer also contains four small, single membrane-spanning subunits: PetG, PetL, PetM and PetN (Figure 1.6A). All major subunits, except the Rieske protein, are encoded in the chloroplast genome, and all minor subunits, except PetM, are products of chloroplast genes in plants and algae. Cyt b_6 binds an atypical haem covalently bound to an invariant cysteine residue, in addition to the two well-known haems (Figure 1.6A) (Stroebel et al., 2003). Cyt *f* contains a *c*-type haem and the Rieske protein ligates

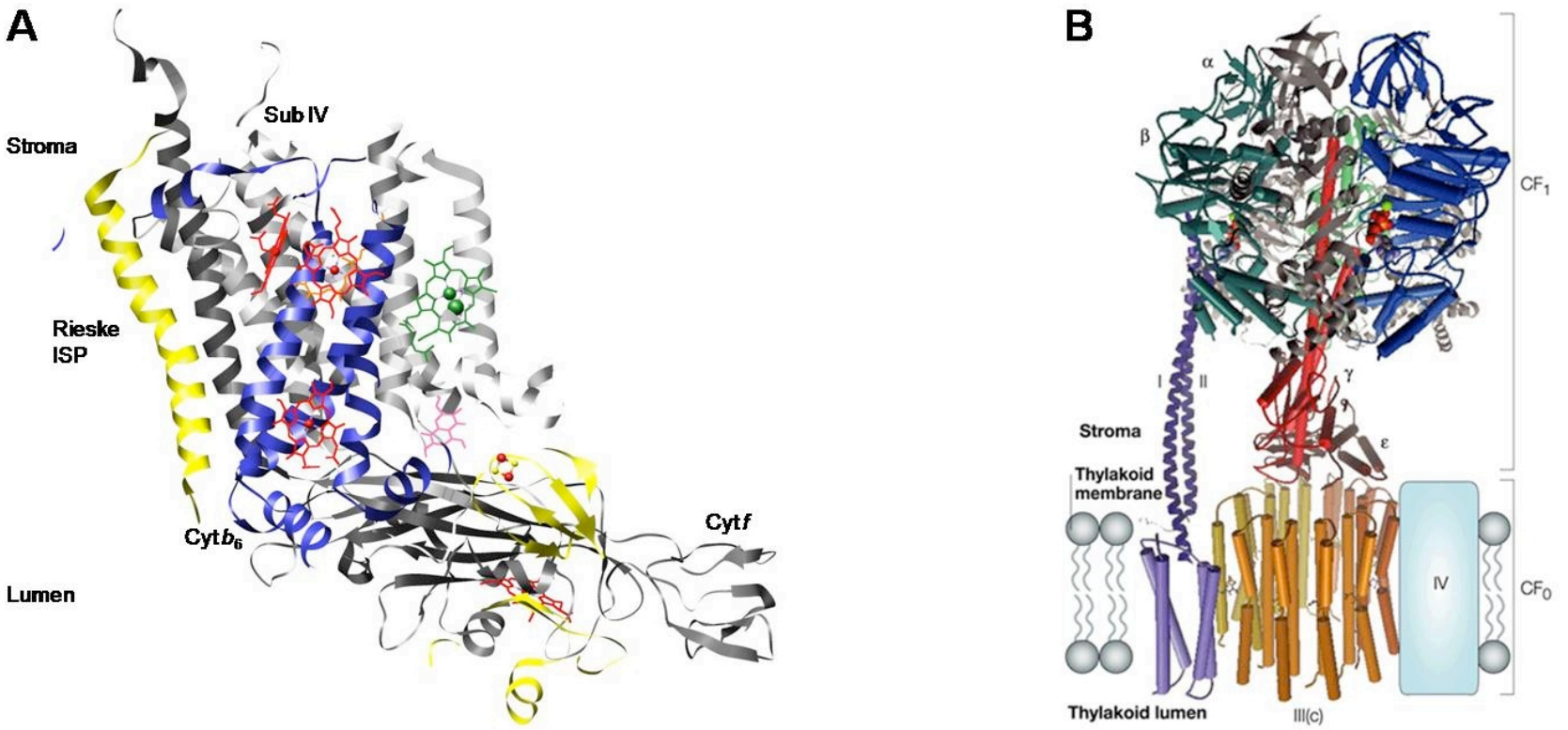


Figure 1.6 Architectures of cytochrome b_6f and ATPase complexes. (**A**) Structure of the Cyt b_6f monomer from *Chlamydomonas reinhardtii*, viewed parallel to the membrane plane. For revealing prosthetic groups and cofactors, some polypeptides and helices of the cyt b_6f are not included in this image. The porphyrin rings of haem are shown in red and the chlorin ring of the chlorophyll in green. Iron-sulphur centres are shown as red and yellow spheres. (**B**) A model for the chloroplast ATPase built around mitochondrial ATPase. α-helices are represented as cylinders. The δ subunit is not included in this image. The C-ring of chloroplast ATPase contains 14 subunits rather than 12 shown in this image. (Figure adapted from (Allen, 2004, Nelson and Ben-Shem, 2004))

two iron-sulphur (2Fe-2S) clusters (Figure 1.6A). Cyt *b6f* also contains one chlorophyll and one carotenoid of uncertain function (Figure 1.6A) (Allen, 2004). Structural elucidation of Cyt b_6f confirmed the notion that it shares a common evolutionary origin with the respiratory cytochrome b-c_1 complex (Allen, 2004).

ATP synthase (ATPase) is a multimeric enzyme complex that drives the formation of ATP from ADP and Pi by using the protonmotive force generated from photosynthetic electron transport (Figure 1.6B). ATPase consists of two distinct multisubunit portions - the membrane embedded CF_o and the membrane extrinsic CF_1. CF_1 is formed by five different polypeptides found in a stoichiometry of $\alpha_3\beta_3\gamma\delta\varepsilon$ (Figure 1.6A) (Abrahams et al., 1994). The CF_o portion consists of four polypeptides (I-IV) in a stoichiometry of 1:1:14:1. The three catalytic sites in the ATPase enzyme are formed by β-subunits of the CF_1. ATPase is a rotary enzyme (Junge et al., 1997). The central c-ring (formed by subunit III) of the CF_o acts together with the γ subunit of the CF_1 as the rotor. The remaining subunits act together as the stator. The CF_o portion of the ATPase acts as proton channel though which protons from the thylkoid lumen flow to the stroma. This flow of protons causes the rotation of the c-ring, and of the γ subunit, which is directly attached to it. The rotation of the asymmetric γ subunit of CF_1 causes a series of conformation changes in the catalytic nucleotide binding sites of the β subunits, which lead to ATP synthesis. Six subunits (α, β, ε, I, III and IV) of the chloroplast ATPase are encoded in the chloroplast genome, while the remaining three subunits (γ, δ and II) are products of nuclear genes.

1.6 The Chloroplast genome and the need for a two-component system

At the beginning of 20th century, the existence of a functional genome in chloroplasts was suggested by studies of non-Mendelian inheritance of variegation in higher plants (reviewed in (Bogorad, 1998). These studies revealed that the genetic determinants of variegation are associated with chloroplasts. This discovery was followed by demonstration of DNA in chloroplast in the 1950s and 60s, culminating

in the demonstration of unique chloroplast DNA sequences and the publication of the first physical map of the chloroplast genome in maize (Bedbrook and Bogorad, 1976). The first complete chloroplast genome sequences, of *Nicotiana tabacum* and *Marchantia polymorpha*, were published in 1986 (Ohyama et al., 1986, Shinozaki et al., 1986). Many chloroplast genomes have been sequenced since, but the miniature genomes of chloroplasts continue to captivate the imagination of scientists working in this area. The average chloroplast genome of higher plants is 120 kbp long and encodes around 120 proteins (Sugiura, 1998), while a typical cyanobacterium, from which the chloroplast evolved, has around 3000 proteins. The coding capacity of chloroplast genome represents only 0.1 % of a typical eukaryotic genome. The size of chloroplast genome is thus miniscule by any standard.

It is believed that the great majority of genes that came with the original cyanobacterial symbiont have been relocated or lost during the endosymbiotic evolution that ensued (Martin et al., 2002). Relocation of genes encoding chloroplast proteins to the nuclear genome is thought to be selectively advantageous (Allen and Raven, 1996) as it reduces the mutation frequency imposed by the redox chemistry operating in chloroplasts. It is also suggested that relocation of chloroplast genes to nucleus prevents the demise of chloroplast genes by Muller's ratchet – a consequence of the accumulation of mutations from asexual propagation of chloroplasts. The majority of the chloroplast proteins today are synthesised in the cytosol as nuclear gene products and imported into chloroplasts as protein precursors (Ellis, 1984). Chloroplast to nuclear gene transfer is ongoing and chloroplast genes are being transferred to nuclear genome at an alarming rate (Huang et al., 2003, Stegemann et al., 2003). These observations, together with the huge cost and complexity of maintaining separate genomes in chloroplasts, pose the question, why are any genes left in the chloroplast?

Various explanations have been put forward to the enigmatic retention of the chloroplast genome (Race et al., 1999). Until recently, one widely-held notion was that some of the chloroplast proteins are too hydrophobic to be imported into

chloroplasts. However, there are some conspicuous counter–examples to this suggestion. The gene encoding the large subunit of the RuBisCO, *rbcL*, with some rare exception in predatory protists, is always chloroplast-encoded. RuBisCO is a stromal enzyme. The hydrophobicity argument therefore fails to account for the chloroplast localization of *rbcL* gene. The light harvesting complex II (LHC II) of PS II, one of the most abundant membrane proteins on earth, is a highly hydrophobic protein with three membrane-spanning helices. The LHC II protein, however, is not encoded in the chloroplast, but is the product of a nuclear gene. Another explanation for plastid genomes was that since organellar genes have non-universal genetic codes, the chloroplast genes could never be translated on cytosolic ribosomes. This suggestion was also disproved by experiments showing successful translation of nuclear-relocated chloroplast genes by cytosolic ribosomes. Some others have argued that, given enough time, all chloroplast genes will eventually end up in nucleus. However, this suggestion would require a random sample of genes remaining in chloroplasts, but the non-random nature of the constant sub-set of chloroplast genes is distinctly at odds with this proposal.

A relatively recent hypothesis, known as CORR (Allen, 1993a, Allen, 2003b), invokes a regulatory reason for the stubbornness of chloroplast genes. This hypothesis suggests that chloroplasts retain genes whose expression is regulated by the activity of the photosynthetic electron transport chain, using a mechanism conserved from their cyanobacterial ancestors. According to CORR, the conserved mechanism that mediates the precise switching on and off of chloroplast genes in response to photosynthetic activity is formed by a two-component system, and this is why chloroplasts need a two-component system (Figure 1.7).

1.7 Chloroplast genomes: reduced to the CORR

The CORR hypothesis (*co*-location for *r*edox *r*egulation) (Allen, 2003b) posits that the small but functional genomes in chloroplasts and mitochondria provide a location for genes and their gene products in the same cellular compartment. This co-location

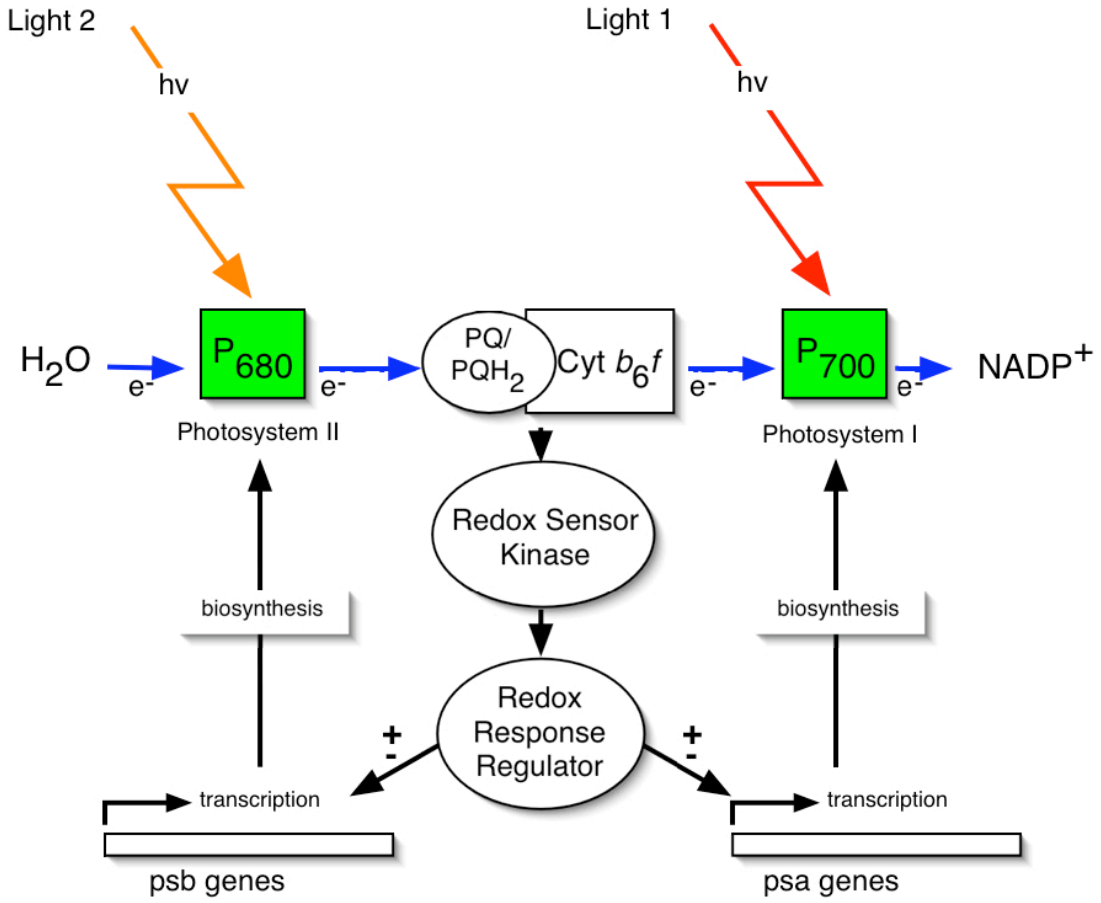


Figure 1.7 The proposed role of a redox-responsive two-component system in light quality mediated transcriptional response in chloroplasts. Light reactions of photosynthesis are represented as electron transport from H_2O to NADP via two photosystems connected by a cytochrome b_6f complex. The redox state of the plastoquinone pool is sensed directly, or indirectly via the cyt b_6f complex, and initiates the transcriptional response. The transcriptional response involves the activation of a DNA binding response regulator protein through phosphorylation, the activated response regulator then regulates the transcription of reaction centre genes for both PS I and PS II. After a switch between light 1 and light 2, the photosystem that becomes light-limiting is up-regulated, while the photosystem that becomes light-saturated down-regulated. (Figure taken from (Puthiyaveetil and Allen, 2008))

of genes and gene products in organelles is selected in evolution in order to place a direct and unconditional regulation on the expression of genes by the redox state of their gene products. Some of the key assumptions of this hypothesis are as follows: bio-energetic organelles retain a constant sub-set of genes that encode for the core-subunits of the electron transfer complexes; the expression of these genes is under the direct control of the redox state of their gene products; and that the mechanism, which ensures redox-control of organellar gene expression, has been retained from the bacterial ancestors of chloroplasts and mitochondria (Allen, 2003b).

Since two-component systems are the major signal transducers in bacteria and some two-component systems are redox-responsive, the conserved mechanism that couples redox signals to gene expression in organelles is predicted to be a two-component system (Figure 1.7). The CORR hypothesis further assumes that those genes that do not encode core-subunits of electron transfer complexes and whose expression need not be redox-regulated can be relocated to the nucleus. This hypothesis also predicts that there is no barrier to the successful import of precursor protein into chloroplasts and mitochondria. Genetic system genes encoded in the bio-energetic organelles, according to CORR hypothesis, are retained for regulation by the primary process of redox chemistry or by secondary signals reporting the current state of gene expression (Allen, 2003b).

Key assumptions and predictions of the CORR hypothesis, since it was first proposed in 1993 (Allen, 1993a), have been borne out with the available data. Comparative analysis of mitochondrial and chloroplast genomes has shown that genes remaining in organelles boil down to two main functional categories: genes encoding structural proteins that work at the core of the redox chemistry, and genes encoding components of the gene expression machinery (Race et al., 1999). The selection pressure to retain only these two functional categories of genes in organelles is further manifested in the observation that organellar genome reduction in parallel and independent lineages has left the same two functional categories of genes remaining in the organelles (Race et al., 1999).

Recent studies have revealed that redox-regulation occurs at all facets of organellar gene expression. Considerable data have accumulated, especially in chloroplasts, to suggest the pervasiveness of redox-regulation in organelles (Pfannschmidt and Liere, 2005). Besides transcriptional control of gene expression, posttranscriptional component of gene regulation has been well established in chloroplasts and mitochondria and in their prokaryotic ancestors. Since both transcription and translation take place in the same cellular compartment in prokaryotes, these two processes are coupled. This means that in prokaryotes, the mRNA is translated while still being synthesized. This coupling ensures rapid and efficient translation of nascent mRNA species, whose half-life in bacteria is a few minutes (Selinger et al., 2003).

While chloroplast mRNAs are relatively long-lived, coupled transcription and translation may still ensure rapid and efficient translation of some mRNAs. However, translation of chloroplast transcripts is mainly regulated by light via a group of nuclear-encoded proteins known as translational activators (Manuell et al., 2004). The light-induced translation of chloroplast mRNAs are, in fact, mediated by changes in the redox state of certain components of the photosynthetic electron transport chain (Manuell et al., 2004). The changes in the redox state of electron carriers regulate the RNA-binding property of translational activators. A well-known example is the regulation of translation of *psbA* mRNA by the ferredoxin-thioredoxin system in *Chlamydomonas* (Danon and Mayfield, 1994). The translational activation of *psbA* mRNA involves the chemical reduction of a translational activator protein by the reduced thioredoxin. This reduction increases the activator's capacity to bind the 5' UTR of *psbA* mRNA, which in turn allows the recruitment of *psbA* mRNA onto polysomes and its subsequent translation.

In chloroplasts, various other post-transcriptional steps such as translational elongation, RNA stability and RNA splicing are subjected to photosynthetic redox regulation (Pfannschmidt and Liere, 2005). As discussed in the next section,

transcriptional regulation has, however, recently emerged as a major theme of redox-regulation in chloroplasts. A related theme, which forms the central subject of research presented in this thesis, is whether chloroplasts retained one or more two-component systems from their cyanobacterial ancestors as the mechanism that permits redox-regulation of their gene transcription (Figure 1.7) (Allen, 1993a).

1.8 Chloroplast gene transcription is regulated by redox signals

Regulation at the level of transcription of DNA to RNA has long been recognized as a principal mechanism for modulating gene expression in biological systems. Gene expression in chloroplasts has been viewed as an exception to the rule of transcriptional control (GoldschmidtClermont, 1998, Gruissem and Tonkyn, 1993). The proponents of this view draw from the observation of relatively long-lived chloroplast mRNA transcripts, and argue that all gene regulation in chloroplasts is post-transcriptional. However, the notion of the non-regulation at the transcriptional level was challenged by a series of studies demonstrating specific light quality induced changes in the transcriptional pattern of chloroplast genes (Fey et al., 2005b, Pfannschmidt et al., 1999a, Pfannschmidt et al., 1999b, Puthiyaveetil and Allen, 2008, Tullberg et al., 2000). The exact nature of the action of light on chloroplast transcription was revealed by the use of various electron transport inhibitors which showed that the light-regulated transcriptional changes of chloroplast genes are, in fact, mediated by alterations in the reduction-oxidation (redox) state of a component of the photosynthetic electron transport chain (Pfannschmidt et al., 1999a). This component is the quinone electron acceptor in chloroplasts known as plastoquinone (PQ) (Figure 1.7).

Redox-regulated chloroplast transcription is part of an acclimatory response in chloroplasts called photosystem stoichiometry adjustments (Chow et al., 1990). As discussed in earlier section, in photosynthesis, two photosystems with different light absorption and action spectra are connected in series to drive electrons from water to $NADP^+$. Their serial connection means that the rate of electron transport between the

two photosystems must be equal. The function of photosystem stoichiometry adjustment is to compensate for any deficiency in energy conversion at either PS I or PS II by increasing the quantity of the photosystem that will otherwise become rate-limiting to overall photosynthesis. For chloroplast transcriptional regulation this means, under conditions in which PS II is selectively excited, that transcription of PS I reaction centre genes is up-regulated so as to increase the number of PS I units. Simultaneously, the transcription of PS II reaction centre genes is suppressed to decrease the number of PS II units. An opposite chloroplast transcriptional response, that is, up-regulation of PS II genes and down-regulation of PS I genes, is seen in light conditions that favour PS I (Pfannschmidt et al., 1999a). An alternative model of photosystem stoichiometry adjustment, which involves only the regulation of PS I amount, has been supported by studies in cyanobacteria (Fujita, 1997), *Chlamydomonas* (Murakami et al., 1997a) and some plant species (Fey et al., 2005b, Pfannschmidt et al., 1999a, Tullberg et al., 2000).

Photosystem stoichiometry adjustment, which operates on a timescale of hours or days, also involves redox-regulated transcriptional regulation of nuclear-encoded photosystem genes (Pfannschmidt et al., 2001, Schutze et al., 2008). As will be discussed later, the rapid transcriptional regulation of chloroplast-encoded core reaction centre proteins has the upper hand in setting the rhythm of the biogenesis of electron transfer complexes in thylakoid membranes. Different plant systems and transcript quantifying methods have been used to demonstrate redox-regulated chloroplast gene transcription. Redox-responsive chloroplast transcriptional regulation has been demonstrated in higher plants such as mustard and pea as well as in the unicellular alga, *Chlamydomonas*, using northern blotting techniques (Pfannschmidt et al., 1999a, Tullberg et al., 2000, Kovacs et al., 2000). Using primer-extension analysis and quantitative PCR (qPCR), the transcriptional regulation of *psaA* gene, which encodes for the PS I reaction centre apoprotein A, has been shown in *Arabidopsis* (Puthiyaveetil and Allen, 2008, Fey et al., 2005b). The latter method has also revealed a greater magnitude of redox-regulation of *psaA* transcription in *Arabidopsis* (Puthiyaveetil and Allen, 2008). The transcription of the

psaA gene has consistently responded in a functionally intelligible way to changes in the redox state of PQ in all of these studies, while the transcription of the *psbA* gene, which encodes for the D1 reaction centre apoprotein of PS II, responded only in some of the plant species tested (Pfannschmidt et al., 1999a).

As shown by run-on transcriptional assays, PQ-responsive transcriptional regulation in chloroplasts originates from transcriptional activation or repression of specific chloroplast genes. The mechanism that connects the redox-state of PQ pool to the chloroplast transcriptional machinery, however, is unknown. Apart from the light quality induced, PQ-mediated transcriptional responses, changes in light fluence rate (light quantity) also affect transcription of chloroplast genes (Baena-Gonzalez et al., 2001) as well as nuclear photosynthetic genes (Lopez-Juez et al., 2007). This light intensity induced transcriptional response is also mediated by changes in the redox state of photosynthetic electron carriers. In higher plants a thiol-mediated increase in the general chloroplast transcriptional activity is known to occur under high light conditions (Baena-Gonzalez et al., 2001). This transcriptional response, signalled by the redox state of glutathione, operates through phosphorylation of the plastid encoded RNA polymerase (PEP) by a serine/threonine kinase known as Plastid Transcription Kinase (PTK).

Photosystem stoichiometry adjustments also occur in cyanobacteria, in response to changes in both quality and quantity of incident light (El Bissati and Kirilovsky, 2001, Fujita, 1997, Murakami et al., 1997b, Hihara et al., 2002, Herranen et al., 2005). The redox state of PQ or of something close to it, is shown to be the signal that initiates the transcriptional response of photosystem genes in cyanobacteria. A redox-responsive two-component system known as RppAB has been implicated in the cyanobacterial photosystem stoichiometry adjustments (Li and Sherman, 2000). This regulatory scheme, however, has been challenged by a recent study (Lopez-Maury et al., 2002), which showed that the RppAB system instead, is involved in nickel sensing. The phenotype observed in the earlier study (Li and Sherman, 2000) was attributed to the cross-talk between different two-component

systems. In cyanobacteria, what is at dispute then is the identity and not the role of two-component system in photosystem stoichiometry adjustments. This then raises the question, have chloroplasts retained one or more of these two-component systems (Figure 1.7) from their cyanobacterial ancestors as on-off switches of chloroplast transcription?

1.9 Too reluctant to relinquish: chloroplast two-component systems in non-green algae

Sequencing of chloroplast genomes is a routine exercise in molecular systematics of plants and algae. Chloroplast DNA sequencing has thus resolved many cladistic disputes. It has also revealed some unexpected genes in chloroplasts, for example subunits of NAD(P) dehydrogenase. Other unexpected genes show sequence similarity to genes for regulatory proteins that were once thought to be confined to bacteria. These are the genes encoding two-component systems. The repertoire of chloroplast-encoded two component systems in non-green algae consists of a sensor kinase and two response regulators. In non-green algae, chloroplast-encoded two component systems have been reported (Duplessis et al., 2007) in disparate taxonomic groups such as rhodophytes, glaucophytes as well as in some other algal groups , such as cryptophytes, haptophytes and raphidophytes, which derived their chloroplasts from secondary endosymbiosis.

As we shall see in the coming chapters, the distribution of chloroplast-encoded sensor kinase and response regulator components in non-green algae is far from uniform. In some of the non-green algae, the sensor kinase and only a single chloroplast response regulator are present, while some have both response regulators present with the sensor. Some of them have no sensor but only response regulator genes in their chloroplasts. A few others have chloroplasts devoid of any two-component system genes. As to the functional role of chloroplast-encoded two-component systems in non-green algae, no definite information exists, except that the

sensor kinase has a putative redox sensor domain and that both of the response regulators belong to the transcription factor family of proteins.

1.10 Whatever happened to chloroplast two-component systems in the green lineage?

The chloroplast genome of green algae and plants, in contrast to that of non-green algae, contain no two-component system genes. The prevalence of chloroplast genes for two-component systems in non-green algae and their curious absence in green algae and plants made some authors (Simpson and Stern, 2002) to wonder "If two-component systems are common in non-greens, what has become of them in the chlorophyte lineage"? Can this apparent absence of chloroplast genes for two-component systems mean the complete loss of these proteins from the chloroplasts of the green lineage? Chloroplast two-component system or any other chloroplast regulatory proteins, according to the CORR hypothesis, belong to the major class of proteins, whose genes need not be redox-regulated and hence can be relocated to the nucleus (Allen, 2003b).

In credence to this view, two-component genes have been identified in the nuclear genome of green algae and plants (Saito, 2001, Thomason and Kay, 2000). The discovery of the ethylene receptor in 1993 is in fact the first report of a two-component system gene in the nuclear genome of plants (Chang et al., 1993). This finding has also the distinction of being the first ever report of a two-component system gene in eukaryotes. Numerous two-component system genes have since been identified in the nuclear genome of green algae and plants. Thus the absence of chloroplast-encoded two-component systems in green algae and plants does not exclude the possibility that these regulatory systems persist in the chloroplast of green algae and plants as products of relocated nuclear genes.

1.11 My PhD research

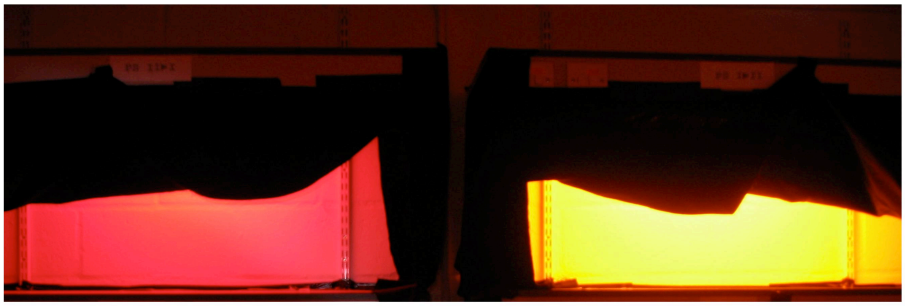
My PhD research was aimed at finding out whether two-component signalling systems exist in chloroplasts of green algae and plants, as predicted by the CORR hypothesis. If chloroplast two-component systems are to be found in the green lineage, my research intends to characterise the function of these proteins in regulating chloroplast gene transcription. This thesis research also explores and analyses the general function, distribution, and evolution of chloroplast two-component systems in photosynthetic eukaryotes as a whole. The model higher plant *Arabidopsis thaliana* was used as the main experimental organism for the research described in this thesis. Sequence analysis tools and basic techniques in molecular biology, biochemistry, genetics and cell biology were used in the research described here.

2.1 Plant material and growth conditions

Arabidopsis thaliana seedlings were grown from seed on soil at 24 ºC at a photon flux density of 100 μE m^{-2} s^{-1} with an 8-hour day and 16-hour dark photoperiod unless otherwise specified. Chlorophyll *a*/*b* ratios were determined from plants that were 2-4 weeks old and grown under two days of light 1, light 2 or white light. For quantifying the transcript accumulation kinetics of the *psaA* gene in response to light quality changes, Wild Type (Col-0) and CSK knockout mutant *Arabidopsis* lines were grown in white light (100 μE m^{-2} s^{-1}; 16-hour day) for 12 days and then transferred to Light 1 or Light 2 cabinets (Figure 2.1) and allowed to acclimate for 4 days. At the end of the 4th day, light 1 was replaced by light 2 and vice versa. Leaves from 2-3 plants were collected for RNA extraction before the light switch and at various time points extending to 32 hours after it.

For studying short-term (≤ 60 minutes) changes in chloroplast transcriptional activity in response to the light 1 to light 2 switch using run-on assays, wild type and CSK knockout mutants were grown from germination in white light (80-100 μE m^{-2} s^{-1}; 16-hour day) for 4 weeks. Plants were then moved to light 1 for 4 days (continuous light, no photoperiod). On the 4th day, light 1 was replaced with light 2. Leaves from 8-9 plants were collected for the run-on assay before switching light 1 to 2 (zero time) and at 30 minutes and 60 minutes after the light switch. For studying long-term (days) changes in the chloroplast transcriptional activity in response to alterations in the quality of incident light, wild type and CSK knockout mutants were grown in white light as described earlier for run-on experiment. Plants were then moved to light 1 or light 2 for 7 days or alternatively plants were moved first to light 1 for 4 days and then shifted to light 2 for 3 days or *vice versa*. Each of the above four light conditions consists of continuous illumination as experiments described earlier. Leaves were collected for run-on assays at the end of each of the above four light regimes.

A



B

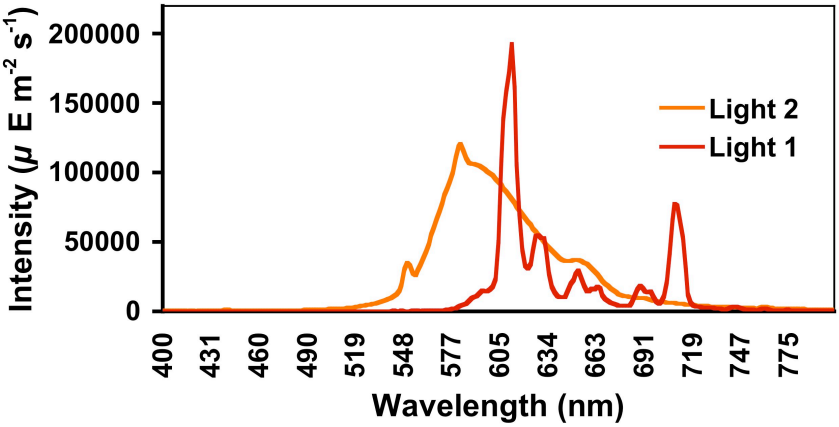


Figure 2.1 Photosystem-specific illumination conditions. (**A**) "Light 1", which selects PS I is rich in far red light components, while "Light 2", which selects PS II, is rich in short wavelength components such as orange. Photograph of the growth cabinets with front curtains partly raised. Photograph taken with a Canon *PowerShot A520* camera with Auto exposure, daylight setting and no colour correction (**B**) Spectral composition of lights 1 and light 2 as measured in similar growth cabinets with the same light sources and optical filters. (Figure 2.1B adapted from (Wagner et al., 2008))

Light 1 (Figure 2.1) was provided by two red fluorescent strip lamps (Osram L 18W/60 Red from Osram GmbH, Hellabrunner Straße 1, 81536 München Germany) each wrapped in red filter (Lee 027 medium red from Lee Filters, Andover, Hants, U.K.). Light 2 (Figure 2.1) was provided by two white fluorescent strip lamps (Osram L 18W/827 Lumillux) each wrapped in orange filter (Lee 105 Orange). The photon flux density at the highest leaves in light 2 was 12 μE m^{-2} s^{-1} and in light 1 was 6 μE m^{-2} s^{-1}. Intensity spectra show extended red and near infra-red components in light 1 compared with light 2 (Figure 2.1B). The selective actions of light 1 and light 2 on photosystem I and photosystem II, respectively, were confirmed by modulated chlorophyll fluorescence and state transition measurements in the growth cabinet (results not shown).

2.2 Isolation of nucleic acids

2.2.1 DNA

2.2.1.1 Genomic DNA

Genomic DNA was isolated from 2-3 leaves of *Arabidopsis* plants with a Qiagen DNeasy Plant mini kit. DNA was eluted in sterile distilled water.

2.2.1.2 Plasmid DNA

Plasmid DNA was isolated from 5-6 ml of overnight cultures of *E. coli* with Qiaprep Spin Miniprep kit from Qiagen. Plasmid DNA was eluted in sterile distilled water.

2.2.2 RNA

Total RNA was isolated from the leaves of *Arabidopsis* plants with a Qiagen RNeasy Plant mini kit. RNA was treated with RNase free DNase (Qiagen) to eliminate possible DNA contamination.

2.3 Quantification of nucleic acid

2.3.1 DNA

DNA concentration was estimated spectrometrically where A_{260} (absorbance at wavelength 260 nm) of 1.0 corresponds to 50 μg/ml DNA. Additionally, concentrations were checked by comparing band intensity of ethidium bromide fluorescence on 1 % agarose gels with that from known quantities of DNA in a DNA ladder (#SM0331, Fermentas).

2.3.2 RNA

RNA concentration was estimated spectrometrically where A_{260} (absorbance at wavelength 260 nm) of 1.0 correspond to 40 μg/ml RNA.

2.4 Polymerase chain reaction (PCR)

2.4.1 Genomic PCR

Primers were purchased from MWG Biotech. Genomic DNA (10-50 ng) was used as template in a final reaction volume of 20 μl, which also contained 0.25 μM final concentration of each of forward and reverse primers, 125 μM final concentration of each dNTP, 0.5 units of *Taq* DNA polymerase and thermopol buffer (NEB). The reaction procedure was: denaturation at 94 ºC for 5 minutes; 35 cycles of: 1 minute at 94 ºC, 30 s at annealing temperature (varied with respect to primer); 1 min/kbp at 72 ºC. Final extension at 72 ºC for 6 minutes.

2.4.2 Reverse Transcriptase PCR (RT-PCR)

Reverse transcriptase PCR was done in a one-step reaction. 200-500 ng RNA template, 0.25 μM final concentration of each of the forward and reverse primers (gene-specific), 0.2 μl of reverse transcriptase mix (Omniscript and Sensiscript, Qiagen) and 10 μl of the 2 X RT master mix (Qiagen) containing HotstarTaq DNA polymerase, dNTPs and the reaction buffer. The final volume of the reaction medium was 20 μl. The reaction procedure was: cDNA synthesis at 50 ºC for 30 minutes; activation of HotstarTaq DNA polymerase at 95 ºC for 15 minutes; 35 cycles of 15 s at 94 ºC followed by 30 s at annealing temperature (varied with respect to primer); 1 min/kbp at 72 ºC. Final extension at 72 ºC for 6 minutes.

2.4.3 Quantitative Real-Time Reverse Transcriptase PCR (qPCR)

Real time quantitative RT PCR was performed with Quantitech SYBR green kit from Qiagen, in a Chromo4 cycler (Bio-Rad). A ~150 bp long sequence was amplified from the *psaA* and the reference gene, *Actin8* transcripts. For amplifying the *psaA* transcript, the forward and reverse primers used were 5' ggcacaagcatctcaggtaa 3' and 5' agcccaaacaatggattcaa 3', respectively, and for the *Actin8* transcript, 5' ttccagcagatgtggatctcta 3' and 5' agaaagaaatgtgatcccgtca 3'. The forward primer for the *Actin8* transcript was designed as flanking an intron-exon boundary thus eliminating the chances of amplifying any contaminating DNA sequences. The optimum annealing temperature for the each primer pair was established by a gradient PCR. The authenticities of the amplificons were confirmed by sequencing the PCR products. Amplification efficiency for each primer pair was calculated by a 16-fold serial dilution of the template and the R^2 value for each primer pair was found to be ≥0.99. A non-template control reaction was done for each primer pair to check whether template-contamination or primer dimers contribute to the fluorescence signals observed. Small fluorescence signals at very late cycle numbers were seen in non-template control reactions for some primer pairs. These signals are likely to have arisen from primer dimers. A non-RT (non-Reverse Transcriptase)

control reaction was also included to check for amplification from any contaminated DNA and it was found that, like non-template controls, a slight fluorescence signal appeared at very late cycle numbers. For measuring the transcription kinetics of light switch samples, RNA was pooled from 2-3 plants and 3 technical replicates were used for each reaction. Expression values were normalized to total RNA. The quantitative real-time PCR technique used here is relative quantification based on comparative C_T method (Cikos et al., 2007).

2.5 Restriction digestion

Restriction digestions were carried out in a final volume of 50 µl, using recommended buffer in conjunction with recommended units of enzyme. Bovine serum albumin (BSA) was added where appropriate. Restriction digestion was carried out at 37 ºC for 2-3 hours unless otherwise specified. Restriction enzymes were purchased from NEB.

2.6 Ligation

Ligations were carried out in 20 µl final volumes at 16 ºC overnight. Reaction tubes contained vector and insert in a ratio of approximately 1:3. T4 DNA ligase (NEB) was used in conjunction with recommended buffer.

2.7 DNA sequencing

1 µg of purified plasmid DNA or 2 ng/µl of PCR amplificons (>300 bp) were sequenced with standard or custom synthesised sequencing primers by the MWG Biotech sequencing service (Germany).

2.8 Sub-cellular localization methods

2.8.1 Transient expression of GFP fusion protein and protein import *in vivo*

Two gene fusions between *Arabidopsis CSK* and the gene encoding the jellyfish green fluorescent protein (GFP) were constructed to investigate the subcellular targeting of CSK–GFP fusion proteins. The construct pCSK1–GFP was generated by amplifying a *CSK* cDNA fragment comprising 54 bp of the 5'UTR and the first 139 codons using Phusion DNA polymerase (New England Biolabs, NJ) and the oligonucleotide primers PK-F (5' cta**ggatcc**gagagtttcagtctcagccacaaagtaa 3'; sequence in bold represents a BamH I cleavage site) and PK-R1 (5' gtat**aggcct**cgagagtactgcgtttggatcaacgat 3'; sequence in bold represents a Stu I cleavage site). The amplified cDNA was digested with BamH I and Stu I and cloned in-frame with the GFP orf in the expression vector as described (Helliwell et al., 2001). The construct pCSK2-GFP contains a cDNA fragment possessing the same 5' end but with the entire *CSK* ORF (without its stop codon) fused in-frame with GFP. This cDNA fragment was amplified using the primer pair PK-F and PK-R2 (5' gtat**aggcct**tgcttcattggcttcagatactgctg 3'). Transcriptional control of the *CSK-GFP* gene fusions was provided by the cauliflower mosaic virus (CaMV) 35S promoter, and transcription termination and polyadenylation signals were provided by the nopaline synthase (*nos*) terminator (Helliwell et al., 2001). Leaves of *Nicotiana tabacum* (tobacco) were bombarded with tungsten particles (0.7 μm) coated with p*CSK*1–*GFP* or p*CSK*2–*GFP* using a Bio-Rad PDS-1000/He particle delivery system, as described (Hibberd et al., 1998). After 24-48 hr, leaf samples of 2 x 2 mm were mounted in water on a glass slide and viewed by scanning laser microscopy (TCS-NT, DMRXA light microscope stand, Leica Microsystems Wetzlar GmbH, Germany). Images of GFP and chlorophyll fluorescence, using an excitation wavelength of 488 nm, were collected through TRITC and FITC filters, respectively.

2.8.2 Chloroplast import of radio-labelled protein precursor

Pea seedlings (*Pisum sativum*, var. Kelvedon Wonder) were grown and then harvested at 9 days old. Chloroplasts were isolated from the leaf tips as described

(Brock et al., 1993). A full-length and a truncated *CSK* cDNA clones were obtained from Genoscope (Paris) and the CSK precursor was synthesised by the method described previously (James et al., 1989) using SP6 RNA-polymerase for cDNA transcription followed by translation in a wheat germ cell-free system in the presence of [^{35}S] methionine. Chloroplast import assays were conducted as before (Mould and Robinson, 1991). Subsequent chloroplast fractionation and protease treatments are done as described before (Di Cola et al., 2005). Controls for chloroplast fractionation were based on published work (Friedman and Keegstra, 1989). An identical import reaction was conducted omitting additional ATP, with the chloroplasts on ice, in the dark. These conditions promote envelope binding but inhibit import.

2.9 Over-expression and purification of GST fusion protein in *E. coli*

A partial cDNA clone (U13211) encoding 450 residues from the carboxy-terminus of CSK, which includes the catalytic domain was obtained from the Arabidopsis Biological Resource Center (ABRC). In order to over-express CSK as a GST fusion protein in *E. coli*, the CSK cDNA was first double digested with EcoRI and NotI and gel purified. The expression vector, pGEX4T2 (Amersham), was similarly double digested and the digestion reaction was cleaned up with the qiagen cleanup kit. The double digested CSK insert and pGEX4T2 vector were ligated with T4 DNA ligase and the ligation products were transformed in XL1 blue cells. Recombinant plasmids were identified by restriction digestion and the identity of CSK insert was confirmed by DNA sequencing using standard pGEX4T2 sequencing primers.

The bacterial strain BL21(DE3) was used for the over-expression of CSK-GST fusion protein. Expression of CSK-GST fusion protein was induced with IPTG at 0.4 mM final concentration when the bacterial cell density reached 0.6 OD. Bacterial cells were harvested three hours after induction and the cells were lysed by sonication. Most of the CSK-GST fusion protein was found in the insoluble fraction, nevertheless some was also found in the soluble fraction. The CSK-GST fusion protein in the soluble fraction was purified by affinity chromatography (Glutathione

Sepharose). In order to prevent aggregation of CSK-GST fusion protein, 10 % (v/v) glycerol, sucrose (1 M final concentration) and arginine hydrochloride (0.4 M final concentration) were present in all buffers and solutions. The GST tagged CSK protein thus obtained was used for the autophosphorylation assay described below.

2.10 Autophosphorylation assay

For the autophosphorylation assay, ~ 10 μg CSK-GST protein was taken in a 25 μl reaction volume containing Tris (50 mM final concentration) pH 7.5, KCl (50 mM final concentration), 10 % (v/v) glycerol, ATP (0.5 mM final concentration), 8 μCi [γ-^{32}P]ATP (3000 Ci/mmol), with or without DTT (2 mM final concentration) and $MgCl_2$ or $MnCl_2$ or $CaCl_2$ at a final concentration of 10 mM. The reactions were incubated at 22 ºC for 60 minutes and terminated by adding sample buffer. The phosphorylated proteins were subjected to SDS-12 % PAGE and blotted onto PVDF membrane (Hybond-P, Amersham). The incorporated phosphate was visualized by autoradiography. The identity of the autophosphorylated CSK-GST fusion protein was confirmed by a monoclonal antibody directed against the GST tag (Novagen) of the fusion protein (results not shown). Standard western blotting procedures (Sambrook et al., 1989) were followed for the immunodetection of the GST tag. Amido black staining of the same membrane was also performed to verify the protein loading. For dephosphorylation of the phosphorylated CSK protein, autophosphorylation of CSK protein was performed as above and 1 μl of Calf Intestine Alkaline Phosphatase (New England Biolabs) was added to the autophosphorylated CSK and incubated at 37 ºC for 30 minutes. Acid/alkali stability of the incorporated phosphate is determined by incubating the membrane in HCl (1 M) or NaOH (3 M) for 2 hours at room temperature. The acid- or alkali-treated membrane was then rinsed with water and subjected to autoradiography.

2.11 Northern analysis

Total RNA was isolated from liquid nitrogen-ground leaf material using TRIzol reagent (Invitrogen). Isolation was carried out according to the manufacturer's instructions. The RNA concentration was determined photometrically. 10-20 μg RNA was taken for each sample and RNA was separated on a 1 % agarose gel under denaturing conditions. RNA was then transferred to nylon membrane using the capillary method and cross-linked with UV light. Hybridization probes were radiolabelled by random priming with Klenow fragment. Purified, PCR generated DNA fragment (50-100 ng; ~500 bp long) was used as the template. Membranes were hybridized overnight at 62.5 ºC, washed, and exposed to the phosphor screen. Signal intensities were calculated with ImageQuant software (Molecular Dynamics). Procedures adopted for the northern analysis were essentially as described in (Sambrook et al., 1989)

2.12 Run-on transcriptional assay of chloroplast genes

Chloroplasts for the run-on assay were prepared by the following method. Leaf material from 9-10 plants (5 weeks old) was homogenized in 50 ml of homogenization buffer at the specified final concentration: sorbitol (0.33 M), HEPES/KOH (50 mM) pH 8.0, $MgCl_2$ (1 mM), Na_2-EDTA (2mM), DTT (0.3 mM) and $NaHCO_3$ (4 mM). Homogenate was first passed through a nylon sieve and then through 4 layers of muslin. Filtrate was centrifuged for 1 minute at 500 RCF. Supernatant from the above step was centrifuged at 2000 RCF for 5 minutes. The supernatant was removed and the pellet resuspended in a small volume of homogenization buffer. Chlorophyll concentration was estimated and chloroplasts equivalent to 30 μg chlorophyll (1-1.5 x 10^6 chloroplasts) were used for the run-on assay.

Chloroplasts run-on assays were performed essentially as described (Mullet and Klein, 1987). Microcentrifuge tubes containing 75 μl of run-on reaction buffer were warmed to 24 ºC and transcription was initiated by adding 25 μl of lysed chloroplasts (equivalent to 30 μg chlorophyll). The reaction medium (final volume

100 μl) contained the following at the specified final concentration: HEPES/KOH, pH 8.0 (50 mM); $MgCl_2$ (10 mM); KAc (25 mM); DTT (10 mM); heparin (0.55 mg/ml); ATP/GTP/CTP (125 μM each); unlabelled UTP (10 μM); [α-^{32}P]UTP (55 μCi; specific activity, 800 Ci/mmol). Run-on transcription was carried out for 7 minutes and the reaction was diluted with 100 μl of H_2O and incubated on ice. 200 μl of Phenol-Chloroform-IAA was added to the reaction mixture and mixed by vigorous shaking, followed by centrifugation at 12000 rpm for 5 minutes at 4 ºC. The aqueous phase (supernatant) was carefully removed and the phenol phase discarded. To the aqueous phase, 5 μl of RNase inhibitor (Fermentas) and 5 μl RNase free DNase (Fermentas) were added and incubated at 37 ºC for 20 minutes. 200 μl of Phenol-Chloroform was added to the reaction mixture and mixed by rigorous shaking. The reaction mixture was then centrifuged at 12000 rpm for 5 minutes at 4 ºC. The supernatant was carefully removed and the phenol phase discarded.

One volume of RNA-extraction buffer (contained the following at the specified final concentration: Tris-Cl, pH 9.0 (1 M); SDS (1% w/v); EDTA (10 mM)) was added to the supernatant. The reaction mixture was centrifuged at 12000 rpm for 5 minutes. 1/10 volume of 3 M NaAc, pH 5.2 and 2 volumes of EtOH were added to the supernatant from the last step. This reaction mixture was incubated in liquid N_2 for 15 minutes and allowed to thaw on ice for another 15 minutes. This incubation was followed by centrifugation at 12000 rpm for 15 minutes at 4 ºC and the supernatant was discarded. The pelleted RNA was washed two times in 70 % EtOH. Each wash consists of a centrifugation at 12000 rpm for 5 minutes. The RNA pellet was then dried under vacuum (speed vac 2 minutes). Alternatively RNA pellet can be air dried for 30 minutes. The RNA pellet was resuspended in 20 μl of sterile H_2O and the RNA solubilized by moderate shaking at room temperature for 30 minutes. This was followed by denaturation of RNA at 65 ºC for 5 minutes. Denatured RNA was incubated on ice for 1 minute before hybridization with denatured DNA probes (immobilized on nylon membrane) essentially as described in (Sambrook et al., 1989). Membranes were then washed and exposed to the phosphor

screen. Signal intensities were calculated with ImageQuant software (Molecular Dynamics) and the expression values were normalized with that of *rrn16*.

Probes for the run-on assay were generated by PCR amplifying 200 - 500 bp fragment from the 5' end of the first exon of the selected genes. The following primer pairs were used for the PCR. PsaA_F, 5' aggcttccacagttttggttt 3', PsaA_R, 5' cccaaacatctgactgcattt 3'; PsaI_F, 5' tctaaaaacatatctttccgtagca 3', PsaI_R, 5' gggaaatgttaatgcatctgg 3'; PsbA_F, 5' cttctgcagctattggattgc 3', PsbA_R, 5' cattttctgtggtttccctga 3'; PsbB_F, 5' gctgacgtcccgtttagaag 3', PsbB_R, 5' tagcaccatgccaaatgtgt 3'; PsbD_F, 5' ttccgtgcttttaacccaac 3', PsbD_R, 5' ccatccaagcacgaatacct 3'; NdhC_F, 5' tatagaaccgatcggggatg 3', NdhC_R, 5' ccttttcgccatgcataaac 3'; NdhH_F, 5' gggttgaaggagttgggatt 3', NdhH_R, 5' tcaaagcccctgctttctaa 3'; RbcL_F, 5' gcgtatgtagcttarccc 3', RbcL_R, 5' tccccctgttaagtagtc 3'; RpoB_F, 5' gcttcgaagatgaacctcttcc 3', RpoB_R, 5' ggagtaaacgaccttctagatgc 3'; ClpP_F, 5' cctggagaaggagatacatcttgg 3', ClpP_R, 5' gtcagcaacagaagcccaag 3'; Rps14_F, 5' tcatttgattcgtcgatcctc 3', Rps14_R, 5' acgtcgatgaagacgtgtagg 3'; TrnE_F, 5' agctagtcataccatttcatttca 3', TrnE_R, 5' cccaggggaagtcgaatc 3'; Rrn16_F, 5' tcatggagagttcgatcctg 3', Rrn16_R, 5' gctttacgcccaatcattcc 3'. PCR fragments were gel purified. For each probe, 50-100 ng of DNA was denatured with 0.5 M NaOH and spotted on a nylon membrane using a slot blotter. DNA probes were then cross-linked to the membrane with UV.

2.13 Quantification of chlorophylls

Chlorophylls were extracted in 80 % acetone and chlorophyll concentration was determined spectroscopically by methods established by Porra *et al* as described in (Pfannschmidt et al., 1999b).

2.14 Sequence analysis

Sequence similarity searches were performed with blastp, tblastn and psiblast programs (Altschul et al., 1990). Sub-cellular localization prediction was carried out with the programs TargetP (Emanuelsson et al., 2000), ChloroP (Emanuelsson et al., 1999), WoLF PSORT (Horton et al., 2007), PCLR (Schein et al., 2001) and Predotar (Small et al., 2004). Domains and motifs were identified using SMART (Schultz et al., 1998) and InterPro databases (Quevillon et al., 2005). Sequence alignment was generated with ClustalW (Larkin et al., 2007) and the alignment was edited with Jalview (Clamp et al., 2004). Secondary structures are predicted with JNet (Cole et al., 2008).

2.15 Phylogenetic tree reconstruction

Multiple alignment of the amino acid sequence corresponding to the kinase domain of CSK and its homologues (as delineated by SMART database) was generated across a representative selection of photosynthetic eukaryotes and cyanobacteria. Sequences were retrieved from both JGI and GENBANK databases. The multiple alignment was generated using CLUSTAL X and adjusted manually using MacClade 4.06. CSK tree was reconstructed from 91 characters. Bayesian phylogeny was generated using Mr. Bayes 3.1 (Ronquist and Huelsenbeck, 2003) from 2,000,000 generations divided between two parallel runs of 1,000,000 each with sampling every 1,000 generations. Although the likelihoods for both trees rapidly reached a plateau, 100 burn-in trees were nevertheless removed from both chains when computing the Bayesian topology. The substitution model was inferred using a mixed model of amino acid substitution, and rate across sites variation was modeled on a discrete gamma distribution approximated using 4 gamma categories and 1 category of invariable sites. Bootstraps were generated using PHYML 2.4.5 (Guindon and Gascuel, 2003) using the WAG substitution model and rate across sites variation modeled on an approximate gamma distribution using 4 gamma categories and one category of invariable sites.

3.1 Introduction

The chloroplast genome of the ancient red alga *Porphyra purpurea*, when sequenced in 1993, had, among other typical chloroplast genes, what the authors called "an intriguing gene." This gene, known as *trsB* (for *t*ranscriptional *r*egulatory *s*ystem), was intriguing because it showed similarity to the then known bacterial sensor kinase EnvZ, and was hitherto unknown in sequenced chloroplast genomes (Reith and Munholland, 1993). TrsB has since been variously known as Dfr in *Gracilaria tenuistipitata*, Tsg1 in *Heterosigma akashiwo* or simply as ycf26 in *Porphyra purpurea* (Duplessis et al., 2007). As more plastid genomes were sequenced, it became apparent that this histidine sensor kinase gene was limited in its phylogenetic distribution: *ycf26* is found only in some non-green algal chloroplasts, and is completely unknown in chloroplasts of green algae and land plants. Moreover, the phylogenetic distribution of *ycf26* in non-green algae is also found to be discontinuous, as it is absent from chloroplasts of red algae such as *Cyanidioschyzon merolae* and from chloroplasts of diatoms such as *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* (Duplessis et al., 2007).

Although chloroplast–encoded sensor kinases are unknown in green algae and land plants, the sequencing of their nuclear genomes has now revealed many genes encoding two-component proteins. In the model higher plant *Arabidopsis thaliana*, there are 54 of such genes and at least 16 of them encode putative sensor kinases (Hwang et al., 2002). If chloroplast sensor kinases, like the majority of chloroplast proteins, can be made in the cytosol and imported into chloroplasts as nuclear gene products (discussed in chapter 1), then their genes will have been relocated, in evolution, to the nuclear genome of green algae and plants.

Chloroplast proteins whose genes have been relocated to the nucleus are synthesized in the cytosol in a precursor form that contains a transit peptide at the N-terminal region. This transit peptide or signal peptide specifies the information for chloroplast location. The amino acid composition of chloroplast transit peptides

shows some common characteristics, which can be of diagnostic value in identifying chloroplast proteins (Bruce, 2001). The N-terminal stretch of ~ 10 amino acids in the chloroplast transit peptide is mostly devoid of charged residues. The region that follows this uncharged stretch is rich in serine or threonine residues. These common sequence features of chloroplast transit peptides have been exploited by various sub-cellular prediction programs, which can predict the chloroplast location with a relatively high degree of success (Richly and Leister, 2004).

To examine the possibility that chloroplasts of green algae and plants have one or more sensor kinases as nuclear gene products, I used a genomic approach. The genomic approach was necessitated by some of the unique properties of histidine kinases, which make their identification by other methods difficult. The instability of phosphohistidine in conventional phosphoamino acid analysis makes its identification by biochemical methods challenging (Klumpp and Krieglstein, 2002). This is in addition to an often-overlooked fact that these proteins are present in very low abundance. The sub-stoichiometric amounts of regulatory proteins such as histidine kinases are further testified by the complete absence of these proteins in the proteomics studies of chloroplasts undertaken so far (Kleffmann et al., 2006).

The genomic approach used in identifying sensor kinases in chloroplasts involved, firstly, finding all sensor kinase genes in the genome and secondly, predicting their sub-cellular location. This approach was applied to the completely sequenced genome of the model higher plant *Arabidopsis thaliana*. In choosing *Arabidopsis*, there were additional considerations such as the availability of sequence analysis tools and other resources developed around its fully sequenced genome. The suitability of *Arabidopsis* for molecular genetics studies and, most importantly, the availability of tagged mutant lines in this model organism (Alonso et al., 2003) would make possible the characterisation of any candidate chloroplast sensor kinase identified in our genomics approach.

3.2 Chloroplast Sensor Kinase in *Arabidoposis thaliana*

Sensor kinases in the *Arabidopsis* genome were first identified by sequence similarity searches. Similarity searches were performed with conserved functional domains of known bacterial histidine kinases such as ArcB (Iuchi and Lin, 1992), RegB (Bauer et al., 2003) and EnvZ (Mizuno et al., 1982). Additionally, cyanobacterial sensor kinases such as RppB (Li and Sherman, 2000) and Hik33 (Hsiao et al., 2004) were used to search with the excellent *Arabidopsis* genome database housed by The Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/). Depending on which histidine kinase protein sequence is used for the similarity search, a set of 12-17 histidine kinases can be retrieved.

In order to ensure that all putative histidine sensor kinases in the *Arabidopsis* genome had been identified by the similarity searches used, additional sequence analyses were performed with secondary databases such as SMART (Schultz et al., 1998), Pfam (Finn et al., 2006) and InterPro (Quevillon et al., 2005).

The SMART database, in addition to storing protein families, domains, motifs and repeats, allows the identification and annotation of known domains and motifs in a protein sequence. Conversely, protein sequences that contain a particular domain or domains can be retrieved from the SMART database. The SMART database recognizes two sub-domains in the conserved kinase domain of sensor histidine kinases: the HisKA domain and the HATPase_c domain. The HisKA domain is the site of autophosphorylation and dimerization, and the HATPase_c domain is the ATP binding domain. The latter domain is found in three major classes of ATP binding proteins; histidine kinases, DNA gyrase B and phytochrome-like ATPases (Dutta and Inouye, 2000). Additionally this domain can be found in heat shock protein HSP90 and DNA mismatch repair proteins (Dutta and Inouye, 2000).

The SMART database identifies 22 distinct proteins with the HisKA domain, 46 proteins with HATPase_c domain and 17 proteins with both HisKA and HATPase_c domains in the *Arabidopsis thaliana* genome database. The large

number of protein sequences with the HATPase_c domain reflects the presence of topoisomerases, heat shock proteins and DNA repair proteins in the retrieved sequences, in addition to histidine kinases.

Putative *Arabidopsis* histidine kinases identified in sequence similarity searches and other methods were then fed into the sub-cellular prediction programs TargetP (Emanuelsson et al., 2000), ChloroP (Emanuelsson et al., 1999), iPSORT (Bannai et al., 2002), WoLF PSORT (Horton et al., 2007), PCLR (Schein et al., 2001) and Predotar (Small et al., 2004). Interestingly, only one histidine kinase, the gene product of the *At1g67840* locus in *Arabidopsis thaliana*, is predicted to be located in chloroplasts. We named this putative sensor kinase of *Arabidopsis* chloroplasts ***C***hloroplast ***S***ensor ***K***inase (CSK) in view of its predicted chloroplast location and its similarity to bacterial sensor histidine kinases.

3.3 Gene and protein sequence features of CSK

Chloroplast Sensor Kinase (CSK) is encoded by the *Arabidopsis thaliana* gene locus, *At1g67840*. Full-length cDNA clones and Expressed Sequence Tags (ESTs) support a gene model in which *CSK* codes for a 611 amino acid long CSK protein. CSK has a predicted molecular weight of 66 kDa, which is further supported by the observation that a full-length CSK cDNA clone can be transcribed and translated into a 66 kDa protein *in vitro* (Chapter 4).

The N-terminal region of the CSK protein shows features of a typical chloroplast transit peptide. No clear transmembrane domains can be predicted in CSK. The absence of predicted transmembrane regions argues against CSK being an intrinsic thylakoid membrane protein. The predicted amino acid sequence of CSK and its homologues from whole genomes or expressed sequence tags, can be aligned with sequences of three canonical bacterial sensor histidine kinases and with the sequence of the ycf26 sensor kinase (Figure 3.1). The three bacterial sensor kinases

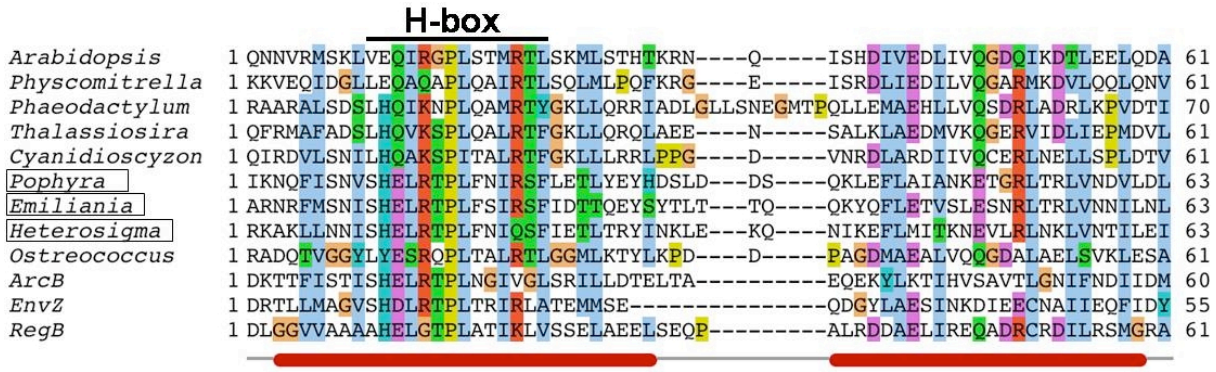


Figure 3.1 CSK shows molecular evolution in its kinase domain. Species names of ycf26 sequences are boxed. The colouring scheme of amino acids is based on their chemical properties and degree of conservation, and is the standard scheme of clustalx (Larkin et al., 2007). (Red, basic; purple, acidic; blue, hydrophobic; green, polar and neutral; brown, glycine; khaki, proline; cyan, histidine and tyrosine). Molecular evolution in the autophosphorylation site of CSK. Sequence corresponding to the HisKA domain (dimerization and phosphoacceptor domain as defined by the SMART database) of CSK and its homologues are aligned with those of ycf26 and with that of three bacterial sensor kinases, EnvZ (Mizuno et al., 1982), ArcB (Iuchi and Lin, 1992) and RegB (Bauer et al., 2003). The sequences shown for HisKA domains correspond to segments between and including the following amino acid positions of the full-length proteins. *A. thaliana*, 305–371; *P. patens*, 295–361; *C. merolae*, 451-517; *P. purpurea*, 417-485; *E. huxleyi*, 386-454; *H. akashiwo*, 140-208; *O. lucimarinus*, 280-340; ArcB (*E. coli*), 282–347; EnvZ (*E. coli*), 233–293; RegB (*R. capsulatus*), 195–261. For *P. tricornutum* and *T. pseudonana*, the sequences are derived from ESTs and partial cDNAs. Predicted secondary structures are shown at the bottom. α-helices are shown as cylinders, beta sheets as thick arrows and loops and turns as lines. The site of autophosphorylation, H-box, located in the first helix of HisKA domain is indicated on the top of the alignment, and the predicted secondary structures are shown beneath them. Alignment was generated with clustalW and edited with Jalview (Clamp et al., 2004). Secondary structures are predicted with JNet (Cole et al., 2008).

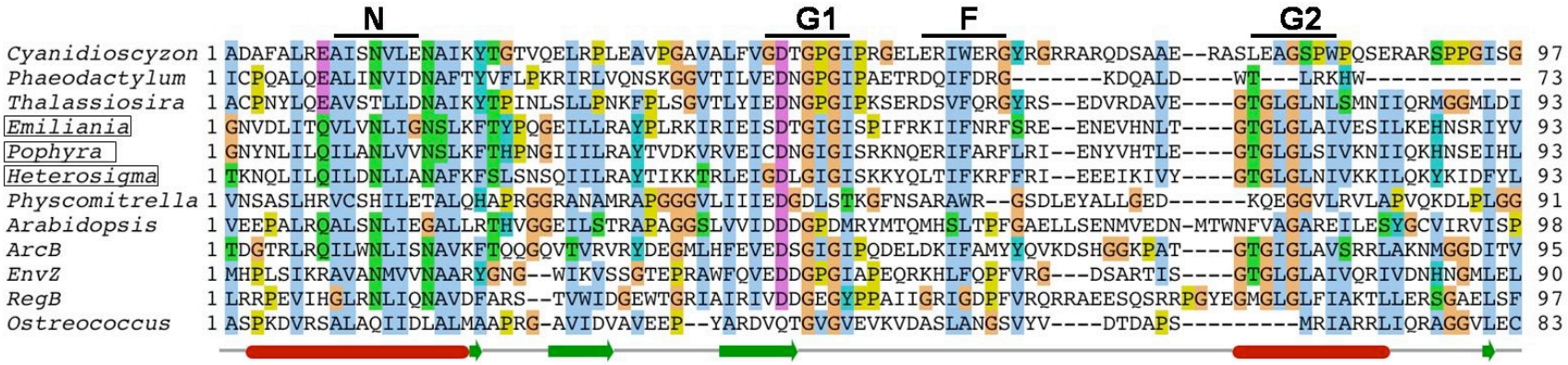


Figure 3.2 The ATP binding domain is conserved in all chloroplast sensor kinases. Species names of ycf26 sequences are boxed. The residue colouring scheme is same as in Figure 3.1. ATP binding domains of representative CSK sequences are aligned with those ycf26 and with that of three canonical histidine kinases, ArcB, RegB and EnvZ. N, G1, F and G2 motifs of the ATP binding domain are shown above the sequences. The sequences shown for the ATP-binding domains correspond to segments between and including the following amino acid positions of the full-length proteins. *C. merolae*, 655-880; *E. huxleyi*, 499-612; *P. purpurea*, 530-654; *H. akashiwo*, 253-369; *P.* patens, 472–584; *A. thaliana*, 478–602; ArcB (*E. coli*), 394–507; EnvZ (*E. coli*), 332–440; RegB (*R. capsulatus*), 303–423; *O. lucimarinus*, 435–534. For *P. tricornutum* and *T. pseudonana*, the sequences are derived from ESTs and partial cDNAs. The predicted secondary structures are shown beneath them. Alignment was generated with clustalW and edited with Jalview (Clamp et al., 2004). Secondary structures are predicted with JNet (Cole et al., 2008).

ArcB, RegB and EnvZ and ycf26 each contain a histidine residue that is the site of autophosphorylation by phosphoryl group transfer from ATP (Figure 3.1). This histidine is retained in CSKs of the red alga *Cyanidioschyzon merolae* and of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* (Figure 3.1). In contrast, *Arabidopsis thaliana* CSK and other plant CSKs contain an homologous "H-box" motif in which the histidine phosphorylation site has been replaced by a glutamic acid residue (Figure 3.1). In further contrast, the histidine phosphorylation site has been replaced by a tyrosine residue in the CSK of the chlorophycean alga *Ostreococcus lucimarinus* (Figure 3.1).

On the other hand, the sequence motifs, N; G1; F; G2; and G3, of the ATP-binding domain of the bacterial sensor kinases and the ycf26 sensor kinase are common to all CSK proteins (Figure 3.2). Further sequence analysis shows a GAF domain, N-terminal to the kinase domain, as the likely sensor domain in CSK and its homologues (Figure 3.3). Secondary databases SMART and InterPro clearly recognize the GAF domain present in algal and moss CSK homologues (Figure 3.3). Even though GAF domains in plant CSK homologues show sequence homology with bacterial GAF domains, they are not clearly identifiable in SMART and InterPro database searches.

3.4 CSK in the DNA microarray database

The availability of the complete genome sequence of *Arabidopsis thaliana* paved the way for large-scale, high throughput gene expression studies in this model organism. Large-scale gene expression studies aided by the DNA microarray technology have resulted in a deluge of expression data in *Arabidopsis*. Various computational tools have been developed in order to store, retrieve and analyse these huge amounts of expression data and most of these data have been made publicly available. It is thus possible that one can analyse the expression pattern of one's favourite gene(s) in these datasets, without actually performing the microarray experiments oneself.

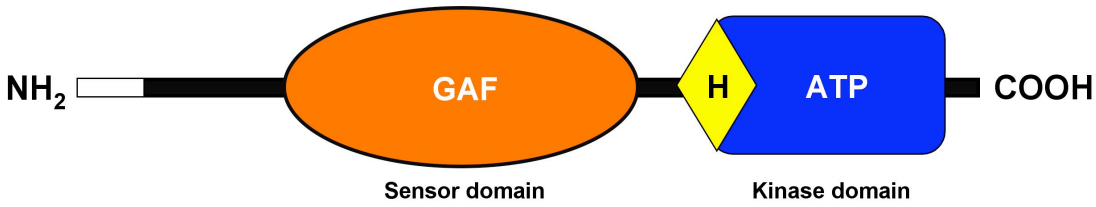


Figure 3.3 Predicted domain architecture of CSK. The amino (NH_2) and carboxy termini (COOH) are indicated. The white rectangle at the amino terminus represents the chloroplast transit peptide. For some CSKs, no GAF sensor domain can be discerned in database searches. Likewise, the conserved histidine autophosphorylation site in the kinase domain (yellow diamond) has been lost in green algal and plant CSKs. "ATP" in the kinase domain denotes the ATP binding domain.

Analysis of the expression of the *CSK* gene in publicly available expression databases showed that the *CSK* is well represented in different datasets. The *Spot history* tool (NASCArrays, http://affymetrix.arabidopsis.info/narrays/spothistory.pl) shows how a gene is expressed over all the experiments in the database. The *spot history* tool reports that the experimental condition "Systemic signalling of irradiance and CO_2 concentration in Arabidopsis" registers the highest expression of *CSK* gene. The lowest expression of *CSK* gene corresponds to "Control of lignification". A similar program called *AtGenExpress Visualization tool* (AVT, http://jsp.weigelworld.org/expviz/expviz.jsp) finds that, under different light conditions tested, only illumination with blue light or far red light causes a relative upregulation of *CSK* gene. The *gene swinger* tool (http://affymetrix.arabidopsis.info/narrays/geneswinger.pl) shows experiments where a given gene is highly variable over all the experiments in the database. The *gene swinger* tool lists cold, drought and oxidative stress as treatments after which CSK expression varies to a great extent. The *expression angler* (Toufighi et al., 2005), finds co-regulated genes in different expression datasets. This program always groups *CSK* gene along with known photosynthetic genes.

3.5 Tissue specific expression of CSK

In order to find out whether the expression of *CSK* gene follows a tissue specific pattern, expression databases were analysed with a publicly available program called *Arabidopsis eFP browser* (Winter et al., 2007). This program creates an "electronic fluorescent pictographic" representation of the expression pattern of the gene of interest. Expression values are derived from datasets named "developmental", "stress", and "hormonal series" of *Arabidopsis* (Winter et al., 2007). It was found that the *CSK* gene is maximally expressed in leaves and other photosynthetic tissues (Figure 3.4). Non-photosynthetic tissues such as roots, mature siliques, seeds and pollen grains showed low or no expression of *CSK* gene (Figure 3.4). This expression pattern of *CSK* was consistent in different datasets. *CSK* expression was seen decreasing with increasing age of the leaf. Amongst different leaf types, leaves

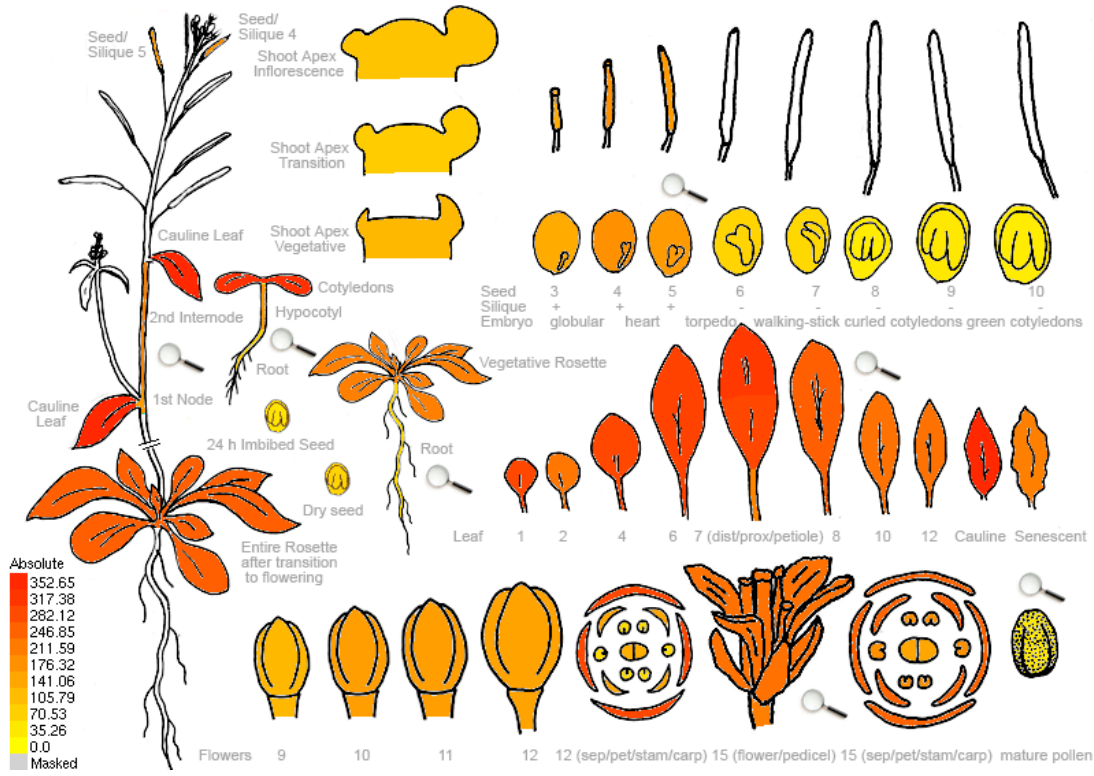


Figure 3.4 Tissue-specific expression of CSK. An electronic fluorescent pictograph showing the expression of the *CSK* gene in different tissue types and at various developmental stages of *Arabidopsis thaliana*. This pictograph was generated using the *Arabidopis eFP browser* (http://www.bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). The microarray dataset used was "developmental series" of *Arabidopisis*. Various tissue types and developmental stages are labelled on the pictograph. Tissue types are coloured according to the expression level of the *CSK* gene. The colour key is given on the left. High expression levels are represented by the red colour. Expression values given in the colour key are fluorescent intensities obtained from the original microarray experiment (developmental series). The "magnifying glass" icon on the image was intended for zooming-in to a tissue-specific dataset from the "output" window of the *Arabidopis eFP browser*.

.

that were borne on the inflorescence axis showed the highest expression of *CSK* (Figure 3.4).

3.6 Discussion

The first reported occurrence of a sensor kinase in chloroplasts refers to the plastid-encoded ycf26 sensor of non-green algae. Although quite prevalent in some non-green algal groups, this sensor kinase is neither universal in non-green algae nor is it reported in green algae and land plants (Duplessis et al., 2007). So far, ycf26 has been found in rhodophytes, *Porphyra purpurea*, *P. yezoensis*, *Cyanidium caldarium*, *Gracilaria tenuistipitata*; in haptophytes, *Emiliania huxleyi* and in raphidophytes, *Heterosigma akashiwo* (Duplessis et al., 2007). Our analysis additionally finds this kinase in the cryptophyte, *Rhodomonas salina*. Though limited in scope for its distribution, this kinase has persisted in genera such as *E. huxleyi*, *H. akashiwo* and *R. salina*, where the chloroplasts have been derived from a secondary symbiotic event involving a red algal ancestor. This points to an overriding functional requirement for this chloroplast sensor in some lineages of non-green algae.

The complete absence of genes encoding sensor kinases or two component systems in general from the chloroplasts of green algae and land plants led many to believe that these proteins may have become extinct in these lineages (Baena-Gonzalez et al., 2001). The possibility of nuclear-encoded sensor kinases being targeted to the chloroplast has been examined before (Forsberg et al., 2001) but none was thought to have a predicted chloroplast targeting sequence (Oelmüller et al., 2001, Wagner and Pfannschmidt, 2006, Lopez-Juez and Pyke, 2005). However, the identification of the Chloroplast Sensor Kinase (CSK) here demonstrates that there is indeed a nuclear-encoded sensor kinase with a chloroplast transit peptide. Other investigators might have missed CSK because its gene is annotated “unknown protein” and not even included as one of 16 obvious histidine kinase genes of *Arabidopsis*. The obvious members of the *Arabidopsis* histidine kinase family include 5 ethylene sensors, 5 phytochromes, 3-5 cytokinin sensors and a putative

osmosensor (Hwang et al., 2002). Poor reliability of some sub-cellular prediction programs may also explain why CSK has been missed in earlier investigations.

In contrast to the ycf26 sensor of non-green algae, the *CSK* gene, has a wide phylogenetic distribution with recognizable homologues in all lineages of green algae and plants, including some red algae and diatoms (see chapter 7). It should be stressed that I identified only one chloroplast sensor kinase in my genomic approach while there could be more than one sensor kinase in chloroplasts. This possibility arises from the fact that not all histidine kinases can be identified with sequence similarity searches. Conventional sequence similarity searches may miss sensor kinases that are highly modified in their primary sequence. The functional domains in these modified sensor kinases may then go undetected in secondary databases, since these currently cannot account for all the sequence evolution in protein families.

In addition to the gene model that encodes for the 611 amino acid long CSK, the TAIR website lists a so called "splice variant" from the same gene locus. This "splice variant" would encode a 445 amino acid long CSK that does not have the C-terminal 166 amino acid residues found in the 611 amino acid long CSK. On closer inspection, I found that this splice variant does not really exist, and it originates from an erroneous electronic annotation based on a single cDNA sequence. The nucleotide sequence of this cDNA clone GSLTFB58ZE06 (as given on the GenBank accession number BX814127), when aligned with the original coding sequence (CD) of CSK, revealed a 4 base pair insertion. This 4 base pair insertion in the cDNA, found at the beginning of the 8th exon, causes an apparent frameshift mutation. The frameshift mutation introduces a premature stop codon in the 8th exon, produing a truncated version of CSK. The 445 amino acid long CSK is thus an artefact of incorrect electronic annotation and not a true splice variant.

Analysis of the protein sequence of CSK revealed both divergent and conserved features. CSK appears to be an unmodified bacterial sensor histidine

kinase in red algae and diatoms as they still retain the conserved histidine residue in the autophosphorylation site (Figure 3.1). On the other hand, the replacement of conserved histidine residue by a tyrosine in green algae and by a glutamate in plants (Figure 3.1) make their CSK a modified histidine kinase. This molecular evolution in the H-box (Figure 3.1), however, does not extend to the ATP binding domain, which appears to be highly conserved in different CSKs (Figure 3.2). Similar molecular evolution in H-box has been observed in plant phytochromes and plant ethylene receptors, which are serine threonine kinases with histidine kinase ancestries (Moussatche and Klee, 2004, Yeh and Lagarias, 1998). The possibility that the kinase activity of CSK has been affected by its molecular evolution is examined in chapter 6.

The predicted GAF domain in CSK might function as its sensor domain (Figure 3.3). GAF is a small-molecule ligand-binding domain first described for vertebrate cGMP specific phosphodiesterase, a cyanobacterial Adenylate cyclase and the bacterial formate hydrogen lyase transcription activator FhlA (Aravind and Ponting, 1997). GAF domains have been known to form the sensor domain in many histidine kinases (Mascher et al., 2006). The importance of the GAF domain in the function of CSK is further discussed in chapter 5.

The expression data clearly indicates that the *CSK* gene is somehow connected with the process of photosynthesis as it responds to factors affecting photosynthetic performance in *Arabidopsis*. The expression data additionally shows that the *CSK* gene is most active in photosynthetic tissues (Figure 3.4) and it is co-regulated with other nuclear-encoded photosynthetic genes. These are clearly the hallmark features of a chloroplast protein and furthermore these insights from the microarray data on the functional role of CSK should inform the selection of experimental conditions that will be used to analyse the phenotype of *CSK* null mutants in *Arabidopsis*.

4.1 Introduction

The Chloroplast Sensor Kinase (CSK), encoded by the *Arabidopsis* gene locus *At1g67840*, is predicted to possess a chloroplast transit peptide (chapter 3). However, the chloroplast location of this protein is to be demonstrated for it to be truly called a Chloroplast Sensor Kinase. An unequivocal demonstration of the sub-cellular localization is also a prerequisite for further functional characterization of this novel protein. During the course of studies directed at localization of CSK, I have collaborated with two research groups in the U.K. The outcome of this collaboration is incorporated into this chapter (Figure 4.1-4.4).

Prof. Tony A Kavanagh and Dr. Christine Newell, working in the research laboratory of Prof. John C Gray at the University of Cambridge, carried out the GFP reporter-based localization studies of CSK (Figure 4.1 and figure 4.2). Mr. Peter Cain, under the supervision of Prof. Colin Robinson at the University of Warwick, carried out the chloroplast import of radiolabelled CSK precursor (Figure 4.3 and figure 4.4). I provided the cDNA clones that were used for generating the radiolabelled CSK precursor. These cDNA clones, originally obtained from the *Genoscope*, Paris (http://www.genoscope.cns.fr/spip/), were extracted from *E. coli* DH5α cells, and the orientation and identity of the CSK insert within the cDNA clones were determined by me in our laboratory at Queen Mary.

4.2 CSK-GFP construct and chloroplast protein import *in vivo*

In order to determine the sub-cellular localization of CSK *in vivo*, the entire coding sequence of CSK (encoding 611 amino acids) was cloned in frame with the reporter gene, jellyfish Green Fluorescent Protein (GFP) (Kavanagh T.A., Cambridge University). Tobacco epidermal cells were then transformed with this CSK-GFP gene construct (Newell C., Cambridge University). After 48 hours, fluorescence microscopy revealed that the transiently expressed GFP is entirely co-located with

chlorophyll fluorescence, showing that the CSK precursor is targeted and imported into chloroplasts (Figure 4.1) (Newell C., Cambridge University).

A second CSK-GFP construct carrying only the first 139 amino acid residues of CSK was also transiently expressed in transformed tobacco epidermal cells. These 139 residues from the N-terminus of CSK are sufficient to target and import the GFP protein into chloroplasts (Figure 4.2), thus confirming the chloroplast transit peptide function of the N-terminal stretch of CSK (Kavanagh T.A. and Newell C., Cambridge University).

4.3 Chloroplast import of radiolabelled CSK precursor

In order to investigate further the location of CSK, two cDNA clones of CSK was obtained (Genoscope, Paris). These clones, GSLTPGH82ZC03 and GSLTFB58ZE06, contain full-length or truncated cDNA of CSK in a pCMV-SPORT6 vector background. These cDNA clones were first digested with the restriction enzyme NcoI to determine the orientation of insert with respect to the SP6 promoter. Both clones were found to have their inserts oriented to be transcribed by the SP6 polymerase (results not shown). The identity of the insert in each cDNA clone was confirmed by DNA sequencing using SP6 promoter primers.

The cDNA clones described above were transcribed and translated *in vitro* in the presence of [^{35}S]-methionine. A chloroplast import assay with the [^{35}S]-methionine-labelled CSK precursor was carried out *in vitro* (Cain P. and Robinson C., Warwick University). The radiolabelled CSK protein is seen to be imported into the chloroplast stroma and, interestingly, the transit peptide is not cleaved off, since the imported CSK has the same molecular mass as the initial translation product (Figure 4.3 and figure 4.4). Cleavage of the transit peptide after import is seen for most chloroplast proteins, therefore a “low energy” binding assay (Friedman and Keegstra, 1989) was carried out in order to ensure that CSK is a genuine chloroplast stromal protein (Cain P. and Robinson C., Warwick University). Under these low

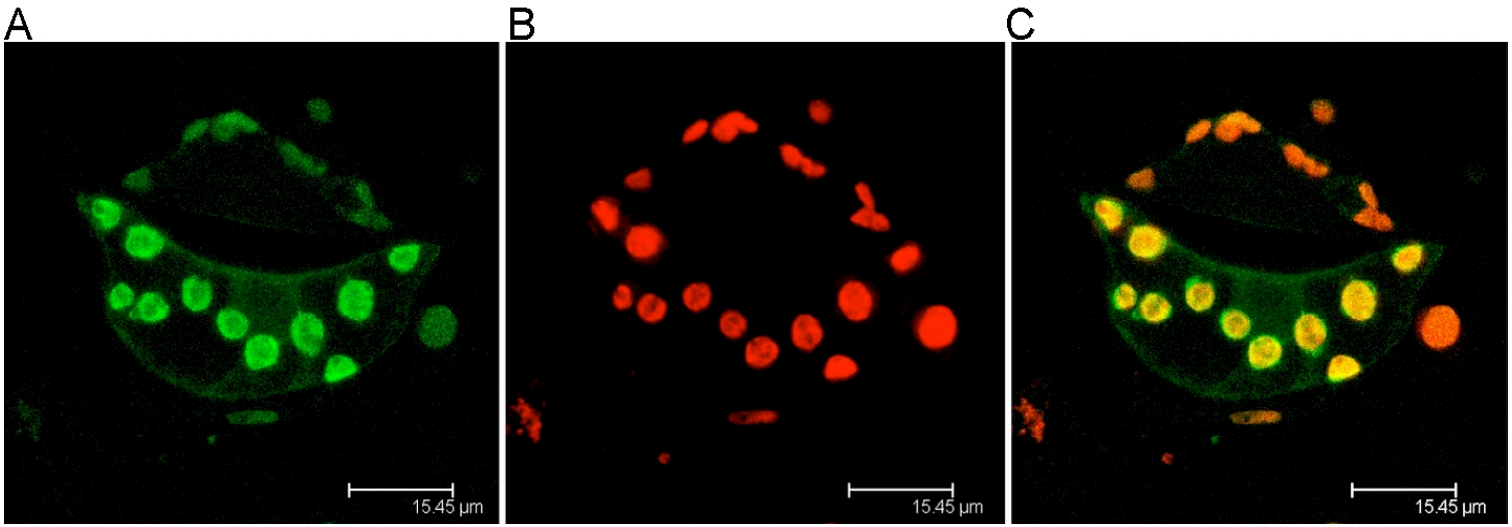


Figure 4.1 A gene construct encoding full-length CSK protein fused with GFP is expressed in the cytosol and the fusion protein is imported into chloroplasts. **(A)** Green fluorescence in one cell of the guard cell pair resulting from its transformation and expression of the CSK-GFP fusion protein. (**B**) Red fluorescence of chlorophyll identifies individual chloroplasts in a pair of stomatal guard cells. (**C**) Overlay of the images in A and B shows that the green fluorescence of the CSK-GFP fusion protein is localized in chloroplasts, where the combined fluorescence of chlorophyll and GFP appears orange. (Image supplied by Kavanagh T.A. Cambridge University)

Figure 4.2 A gene construct encoding a fusion protein of the first 139 amino acids of CSK and GFP is expressed in the cytosol and the fusion protein is imported into chloroplasts. (**A**) Green fluorescence in one the leaf epidermal cells resulting from its transformation and expression of the CSK-GFP fusion protein. (**B**) Red fluorescence of chlorophyll identifies individual chloroplasts in epidermal cells. (**C**) Overlay of the images in A and B shows that the green fluorescence of the CSK-GFP fusion protein is localized in chloroplasts, where the combined fluorescence of chlorophyll and GFP appears orange. (Image supplied by Kavanagh T.A, Cambridge University)

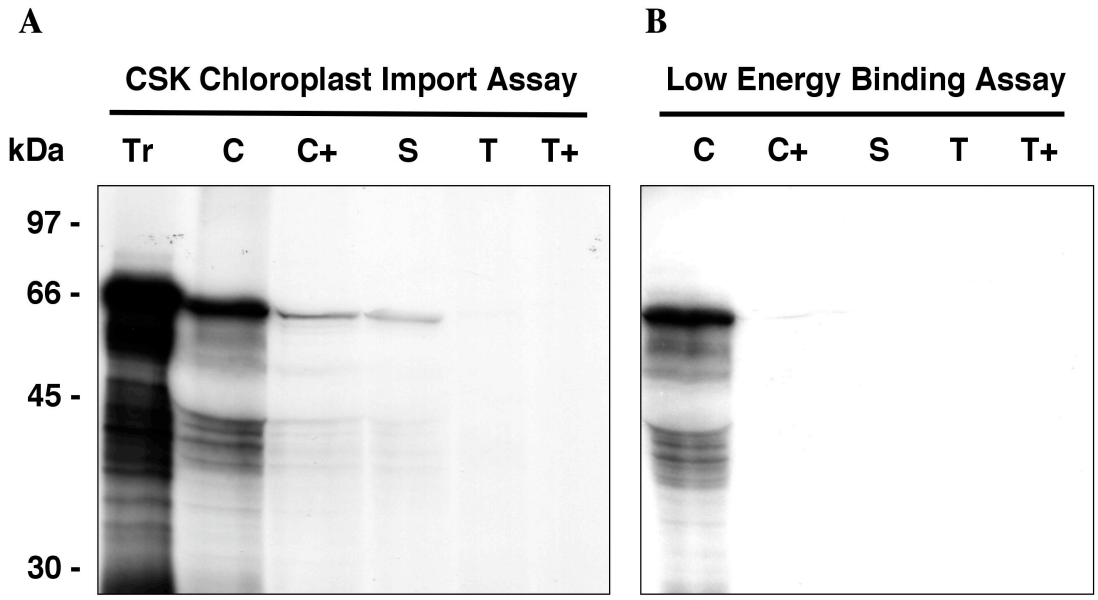


Figure 4.3 Full-length CSK protein (ca. 66 kDa) transcribed and translated *in vitro* from the cDNA clone GSLTPGH82ZC03 is targeted into the chloroplast stroma. (**A**) CSK is a stromal protein with an unprocessed transit peptide. A chloroplast import assay with the [^{35}S]-methionine-labelled CSK precursor showed that the radio-labelled CSK protein is imported into the chloroplast stroma. The protease, thermolysin digests proteins bound on the outer envelope of chloroplasts but this enzyme cannot penetrate into chloroplasts so that the internal proteins are resistant to digestion. Thermolysin digested chloroplasts (in lane C+) shows that the CSK precursor protein is protected from thermolysin digestion on account of being internal to the envelope membrane (being imported into chloroplasts). The transit peptide is not cleaved off because the imported CSK has the same molecular mass as its precursor. (**B**) CSK is a genuine chloroplast stromal protein. A 'low energy' binding assay showed that under low energy conditions, the radio-labelled CSK precursor binds to the chloroplast envelope, ready for import, but remain sensitive to added thermolysin (lane C+). The positions of molecular weight markers are indicated on the left. Tr, translation products; C, chloroplasts; C+, thermolysin digested chloroplasts; S, stroma; T, thylakoids; T+, trypsin digested thylakoids. (Figure supplied by Peter Cain, Warwick University)

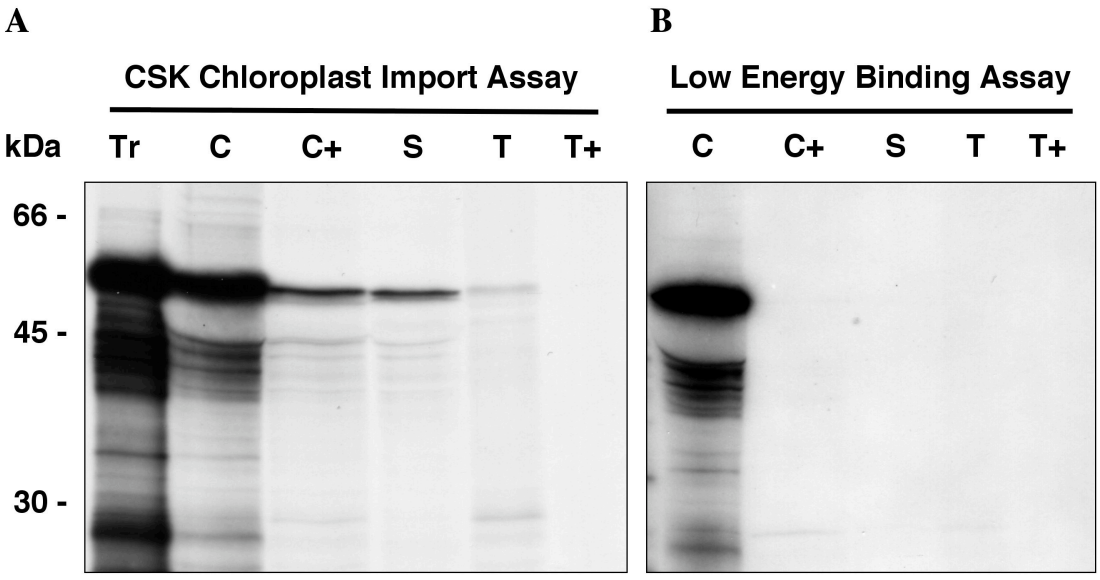


Figure 4.4 A truncated CSK protein (ca. 55 kDa) transcribed and translated *in vitro* from the cDNA clone GSLTFB58ZE06 is targeted into the chloroplast stroma. Experimental conditions and inferences contained in this figure are same as Figure 4.3. (Figure supplied by Peter Cain, Warwick University).

energy conditions – low temperature, without added ATP, and in darkness – chloroplasts are known to bind protein precursors ready for import (Friedman and Keegstra, 1989). This is seen to be the case for CSK where the precursor protein associates with chloroplasts, but remains sensitive to added protease (Figure 4.3 and figure 4.4) (Cain P. and Robinson C., Warwick University).

4.4 Discussion

Chloroplast import of GFP-tagged CSK protein *in vivo* (Figure 4.1 and figure 4.2) and import of [^{35}S]-methionine-labelled CSK precursor *in vitro* (Figure 4.3 and figure 4.4) independently demonstrate the chloroplast localization of CSK. Chloroplast import of radio-labelled CSK precursor additionally provides the information that the CSK is localized in the chloroplast stroma. This import assay further revealed the interesting possibility that the mature CSK protein retains the transit peptide after import. For most chloroplast proteins, the transit peptide is usually cleaved off after import so that the mature protein has slightly lower molecular weight than the unprocessed precursor protein. Some chloroplast proteins nevertheless have unprocessed transit peptides. The nuclear-encoded chlorophyll *a*/*b* binding protein CP29 of *Chlamydomonas* is an example of a chloroplast protein with an unprocessed transit peptide (Turkina et al., 2004). Further studies must examine the possibility that the CSK precursor is indeed processed after import into chloroplasts, but that the transit peptide is too small (Figure 4.5) to be discriminated after cleavage on a SDS-PAGE gel.

If CSK really possesses an unprocessed transit peptide, what function could it serve after import into chloroplasts? In the *Chlamydomonas* CP29 protein the unprocessed transit peptide is a site of regulatory phosphorylation and acetylation (Turkina et al., 2004). It remains to be seen whether the unprocessed serine/threonine rich transit peptide (Figure 4.5) will act as an autophosphorylation site in green algal and plant CSKs as they have lost their conserved histidine autophosphorylation site in evolution. Plant phytochromes, which have similarly lost the conserved histidine

Figure 4.5 Putative chloroplast transit peptide-containing N-terminal extensions of selected CSK orthologues. Predicted CSK N-terminal amino acid sequences from two angiosperms, a moss, a green alga and a red alga are shown. The amino- (NH_2) and carboxy- (COOH) termini are indicated. Sequences in bold and coloured red are regions of CSK proteins, which have homologous regions in the cyanobacterial orthologue of CSK, hik2. Sequences coloured in black shows no homology to hik2 protein and include the predicted chloroplast transit peptide. The predicted cleavage sites of chloroplast transit peptides (ChloroP, (Emanuelsson et al., 1999)) are indicated by the blue arrowhead above the sequence. The sequences shown for the N-terminal extensions correspond to segments between and including the following amino acid positions of the full-length CSK proteins. *A. thaliana*, 1-129; *V. vinifera*, 1-132; *P. patens*, 1-127; *O. lucimarinus*, 1-102; *C. merolae*, 1-176.

residue, autophosphorylate at their serine rich N-terminus (Yeh and Lagarias, 1998). Studying the import of CSK in the red alga *Cyanidioschyzon merolae* and in the diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* can further test the possibility of an N-terminal phosphorylation site. If the unprocessed transit peptide serves as an autophosphorylation site, CSKs in these non-green algae would have their transit peptide (Figure 4.5) processed as they still retain the conserved histidine autophosphorylation site.

It may be also possible that the unprocessed transit peptide is somehow involved in assisting the function of CSK in chloroplasts. One could envisage, for example, function in facilitating transient thylakoid membrane association of CSK; for sensing the redox signals originating from the thylakoid membrane; and for functional associations with components in the CSK signal transduction chain. For the first two possibilities, there should be predictable differences in the import properties of CSK from non-green algae and plants, as they have a different thylakoid membrane composition in terms of light harvesting antenna and other peripheral components.

Stromal location of CSK should be further confirmed with additional methods. The possibility of CSK being a loosely bound thylakoid extrinsic membrane, easily detached in the presence of mild detergents used in experimental conditions, should be examined. Sensitive anti-CSK antibodies could be used to confirm the sub-chloroplast location of CSK. Epitope-tagging of CSK and its over-expression might facilitate co-isolation of CSK with its associating partners by immunoprecipitation.

5.1 Introduction

Having established the chloroplast location of CSK, it is time to ask what functional role does this novel two-component sensor kinase serve in chloroplasts? The general functional properties of two-component systems act as a useful guide for exploring the functional role of CSK in chloroplasts. Sensor kinases of two-component systems sense specific environmental cues and trigger appropriate responses via their cognate response regulator proteins. Since the latter are often DNA-binding transcription factors, the physiological responses mediated by two-component systems usually involve regulation at the transcriptional level (Robinson et al., 2000).

From the general properties of two-component systems, it can thus be surmised that CSK forms the molecular circuitry of transcriptional control in chloroplasts. Is this inferred transcriptional regulatory role of CSK congruent with its molecular evolution? It is seen that CSK, except in red algae and diatoms, appears to be a modified sensor histidine kinase. The conserved histidine phosphorylation site has been replaced in CSKs of green algae and plants and this modification is likely to affect the way in which they transduce signals. A similar molecular evolution in plant phytochromes and plant ethylene receptors, however, does not appear to have an effect on the operation of these important proteins. The underlying mechanism of the physiological responses mediated by phytochromes and ethylene receptors is transcriptional in nature in both eukaryotes and prokaryotes, despite them having been modified in eukaryotes (Quail, 1991, Alonso and Stepanova, 2004).

As detailed in the chapter 1, regulatory signals that modulate chloroplast gene expression are now known, notably as changes in the redox state of key components of energy transduction such as that of plastoquinone (PQ) (Pfannschmidt et al., 1999a, Puthiyaveetil and Allen, 2008, Tullberg et al., 2000). But the mechanism that connects photosynthetic electron transport to gene transcription in chloroplasts remains unknown. A CSK-like sensor kinase thus fits completely in the transcriptional regulatory scheme of chloroplasts, which requires precise sensing and

response to changes in photosynthetic activity.

Since the two photosystems, which are connected in series in oxygenic photosynthesis, have different absorption and action spectra, it is possible experimentally to select the spectral composition of light in order to favour absorption either by photosystem I (with "light 1") or photosystem II (with "light 2"). The preferential excitation of photosystems with light 1 or light 2, introduces imbalance in the electron transport chain and affect the redox state of inter-photosystem electron carriers especially that of PQ. By using these artificial illumination conditions (Figure 2.1, chapter 2, Materials and Methods), chloroplast transcriptional responses can be studied in different plant systems.

Functional characterisation can also depend upon the availability of knockout mutants for the *CSK* gene. As discussed in chapter 3, the selection of *Arabidopsis* in this regard is benefited by the extensive collection of T-DNA tagged gene knockout lines available in this model organism. For elucidating the functional role of CSK, two independent T-DNA lines harbouring insertions in two different positions of the *CSK* gene locus were obtained from the Arabidopsis Biological Resource Center (ABRC). The homozygosity of the T-DNA insertion and its effect on completely shutting down the expression of *CSK* gene has been verified (Appendix) before using these T-DNA lines in the phenotypic characterisation described below.

This chapter on the functional characterization of CSK partly incorporates the outcome of experiments done in collaboration with Lars Dietzel and Dr. Thomas Pfannschmidt at the Friedrich-Schiller-Universität, Germany. As part of this collaboration, I undertook two separate research visits to the laboratory of Dr. Thomas Pfannschmidt at the Friedrich-Schiller-Universität. Data presented in Figure 5.2-5.8 specifically originate from this joint research and are preliminary.

5.2 CSK links photosynthetic electron transport to gene expression in chloroplasts

In order to determine whether CSK is a regulator of transcription in chloroplasts, we examined the transcriptional response of the chloroplast *psaA* gene to changes in light quality, and then compared the transcriptional response in wild-type *Arabidopsis thaliana* with that in two CSK T-DNA insertion lines. It has been demonstrated in different plant species that transcription of the *psaA* gene, which encodes the photosystem I reaction centre apoprotein A, responds robustly and in a functionally intelligible way to changes in the redox state of the electron carrier PQ (Pfannschmidt et al., 1999a, Puthiyaveetil and Allen, 2008, Tullberg et al., 2000).

Figure 5.1 shows the complex kinetics of *psaA* transcript accumulation in wild-type *Arabidopsis* plants. Figure 5.1 also shows clear changes in these kinetics in the two *Arabidopsis* CSK-null mutants after shifts in the spectral quality of incident light that change the redox state of plastoquinone as previously reported (Puthiyaveetil and Allen, 2008). In both mutants tested, the transcriptional response of CSK-mutant plants differs from that of wild-type plants. After acclimating plants in light 1 (favouring photosystem I) condition for four days, light 1 is replaced by light 2 (favouring photosystem II), transcripts of the chloroplast *psaA* gene for the photosystem I reaction centre protein PS I-A accumulate up to 11-fold in 26 hours for the Wild Type (Figure 5.1A), which is the normal functional response of up-regulation of genes for photosystem I (Pfannschmidt et al., 1999a). However, CSK-mutant plants show only a 5-fold increase in *psaA* gene transcription under the same conditions (Figure 5.1A). The reverse light switch, from light 2 to light 1, produces a 2.5-fold decrease in *psaA* expression in the wild type, as the reaction centre of photosystem I is repressed (Figure 5.1B). In contrast, *psaA* transcript quantity does not decrease under the same conditions in CSK-mutants (Figure 5.1B). Instead, the sign of the response is reversed, and *psaA* transcript quantity increases for 8 hours, eventually falling between 26 and 32 hours (Figure 5.1B). As discussed in chapter 1,

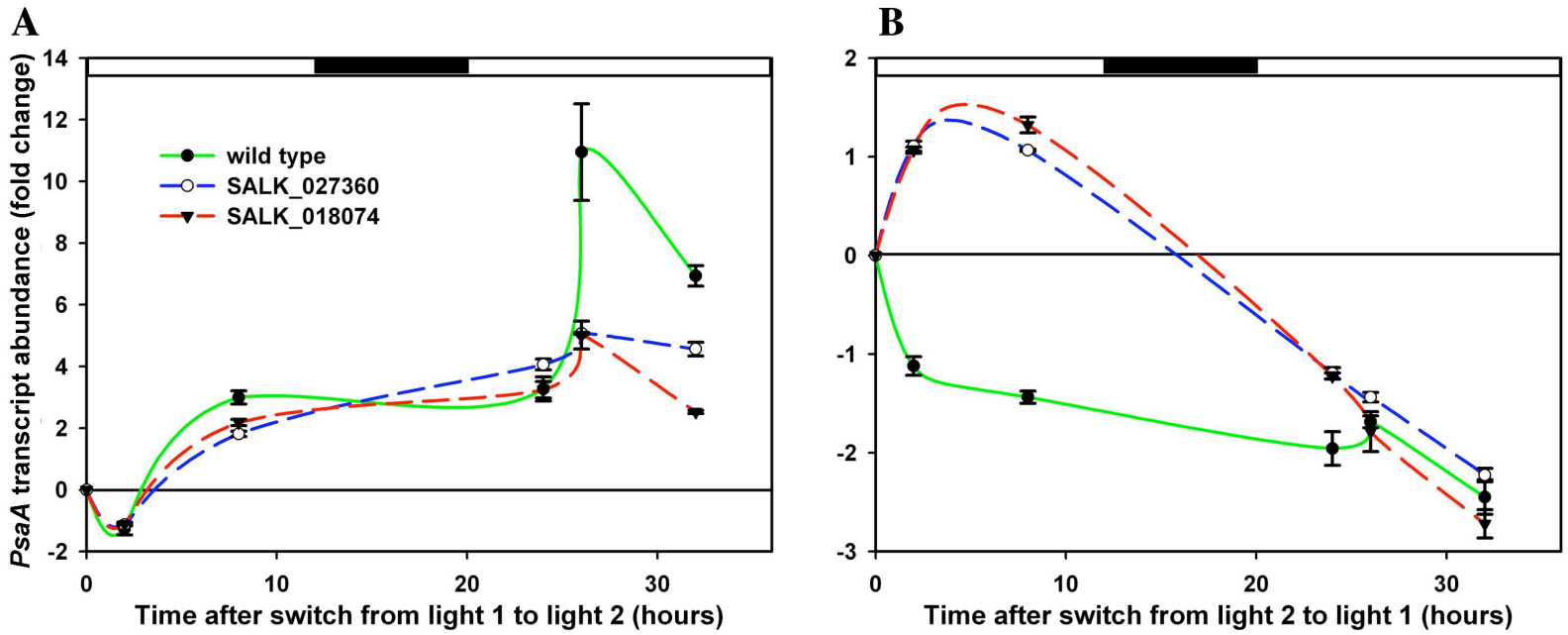


Figure 5.1 Inactivation of the *CSK* gene in *Arabidopsis* results in the loss of photosynthetic control of chloroplast gene transcription. The *psaA* gene transcription kinetics in Wild Type and CSK knock-out mutants, SALK_027360 and SALK_018074, as quantified with qPCR. Changes in gene expression are shown as *psaA* transcript abundance (fold change) plotted against time. Experimental conditions are replacement of light 1 with light 2 and of light 2 with light 1. The time point at which the lights are switched is taken as zero time. The fold change is calculated by taking the expression value (C_T value) of the sample at zero time (before the light switch) as the baseline (calibrator). The difference between expression values of the calibrator (sample at zero time, before the light switch) and of samples at time points, 2, 8, 24, 26 and 32 hours after the light switch were calculated. A difference of 1 between C_T values of the calibrator and of sample at 2 hours, for example, corresponds to 2-fold change (assuming the PCR amplification is logarithmic, 2^n, where "n" is the difference in C_T values). Thus the fold change for sample at zero time is taken as zero fold. Error bars represent ±SE from three technical replicates. An eight-hour dark period is shown as the shaded rectangle on the X-axis.

normal *psaA* transcriptional control is part of an acclimatory response in chloroplasts called photosystem stoichiometry adjustment. The function of photosystem stoichiometry adjustment is to compensate for any deficiency in energy conversion at either photosystem I or photosystem II by increasing the quantity of the photosystem that will otherwise become rate-limiting to overall photosynthesis. Since chlorophyll *b* is located largely in chloroplast photosystem II, the chlorophyll *a*/*b* ratio is a measure of the stoichiometry of photosystem I to photosystem II (Hill and Bendall, 1960). Table 5.1 shows Chl *a*/*b* ratios in wild type and CSK null mutant plants after they were grown in 2 days of white light or light 1 or light 2. The lack of a functional photosystem stoichiometry adjustment, as revealed by the aberrant *psaA* transcriptional response, is further reflected as an inability of the CSK mutants to control Chl *a*/*b* ratio in response to changes between light 1 and light 2 (Table 5.1).

Table 5.1 Chlorophyll *a*/*b* ratios in white-light-grown, light 2-grown and light 1-grown wild type and CSK mutant plants

	Grown in white light	Grown under Light 2	Grown under Light 1
Wild type Chl *a*/*b* ratio	2.27 ± 0.13	3.05 ± 0.03	1.55 ± 0.09
CSK mutant Chl *a*/*b* ratio	2.32 ± 0.08	2.54 ± 0.05	2.10 ± 0.12

± Standard error from three independent measurements

5.3 CSK mediated gene regulation in chloroplasts involves differential promoter usage

Since steady-state mRNA levels reflect both synthesis and degradation of transcripts, it is important to show that the altered transcript accumulation kinetics seen in CSK

null mutants results from differences in transcriptional activation/repression rather than from relative stability of transcripts. Run-on transcription assays (chapter 2, Material and methods) of selected chloroplast genes were performed to assess the kinetics of transcriptional activity in CSK null mutants in response to short-term changes in illumination conditions that affect the redox state of plastoquinone. Chloroplast run-on transcription assay is a technique to look at chloroplast genes being transcribed at a specific time point. Chloroplasts isolated from leaf tissues are first lysed. The lysed chloroplasts contain transcription machinery (RNA polymerase) stalled on the chloroplast DNA template due to acute shortage of ribonucleotides. Transcription can be started up again *in vitro* with the addition of nucleotides. By radio-labelling nucleotides, one can observe which mRNA species are synthesised from the stalled chloroplast gene promoters and thus which chloroplast genes were being actively transcribed at the time chloroplasts were isolated. By comparing the amount of radio-labelled nascent mRNA species hybridized to gene-specific probes in different chloroplast samples taken at different time points, one can assess the relative transcriptional activity of specific chloroplast genes in chloroplast samples of interest.

Plants were first acclimated in light 1 condition for four days and at the end of the 4th day, light 1 was replaced with light 2. Chloroplasts for run-on assays were isolated from 10-12 plants harvested before the light switch (in light 1, zero time), and after 30 minutes, and 60 minutes into the light switch (light 2). Chloroplasts were lysed by pipetting up and down for 10-15 times and run-on transcription assays were carried out for 7 minutes at 24 ºC. The [α-^{32}P]UTP-labelled chloroplast transcripts were hybridized to specific DNA probes immobilised on nylon membrane. The amount of [α-^{32}P]UTP incorporated was calculated by autoradiography. Transcriptional activity of the following chloroplast genes was assayed. *PsaA* encoding the PS I-A apoprotein; *psaI*, PS I-I subunit; *psbA*, D1 reaction centre apoprotein of PS II; *psbB*, CP47 protein of PS II; *psbD*, D2 reaction centre apoprotein of PS II; *ndhC*, D3 subunit of NAD(P)H dehydrogenase; *rbcL*, RuBisCO large subunit and *rrn16* encoding the 16S rRNA.

It is seen that in the wild type, the transcriptional activity of the *psaA* gene decreases as an immediate response to changes in illumination from light 1 to light 2 (Figure 5.2). But at 60 minutes of the light switch, the *psaA* transcriptional activity shows an increase over the activity seen at 30 minutes of the light switch (Figure 5.2). In CSK mutant, the increase of the *psaA* transcriptional activity seen in the wild type at 60 minutes of the light switch is lacking, instead a further decrease from that present at 30 minutes is found (Figure 5.2). The transcriptional activity of the *psbA* gene in response to the light 1 to 2 switch similarly differs between wild type and the CSK mutant (Figure 5.2). In the wild type, *psbA* activity shows a decrease at 30 minutes after the light switch, but at 60 minutes the *psbA* activity returns to the same level as that present before the light switch (zero time) (Figure 5.2). In CSK mutant, the *psbA* activity does not return to the same level as that seen at zero time, and continues to decrease (Figure 5.2).

In the CSK mutant, transcriptional rates of photosystem genes *PsaI*, and *psbB* differed from that of wild type in response to light switch and resembled the pattern seen for *psaA* and *psbA* respectively (Figure 5.2). Transcriptional rate of the photosystem gene *psbD* showed a similar behaviour in both CSK mutant and wild type (Figure 5.2). However in CSK mutant, transcriptional rate at 60 minutes did not reach the same level as that of the wild type (Figure 5.2). In the CSK mutant, non-photosystem genes *ndhC* and *rbcL* also showed altered transcriptional activity in response to the light switch when compared with the wild type (Figure 5.2).

In order to determine the total transcriptional activity of chloroplasts in the above short-term light shift experiment for wild type and for the CSK mutant, a small aliquot of the reaction product from each of the run-on assays was spotted onto a Whatman-silica filter paper and the [α-^{32}P]UTP incorporation was quantified. The total [α-^{32}P]ATP incorporation at zero time (before the light 1 to 2 switch) was similar in both wild type and CSK mutant, but at time points 30 and 60 minutes after the light switch, the total [α-^{32}P]UTP incorporation was notably higher in CSK

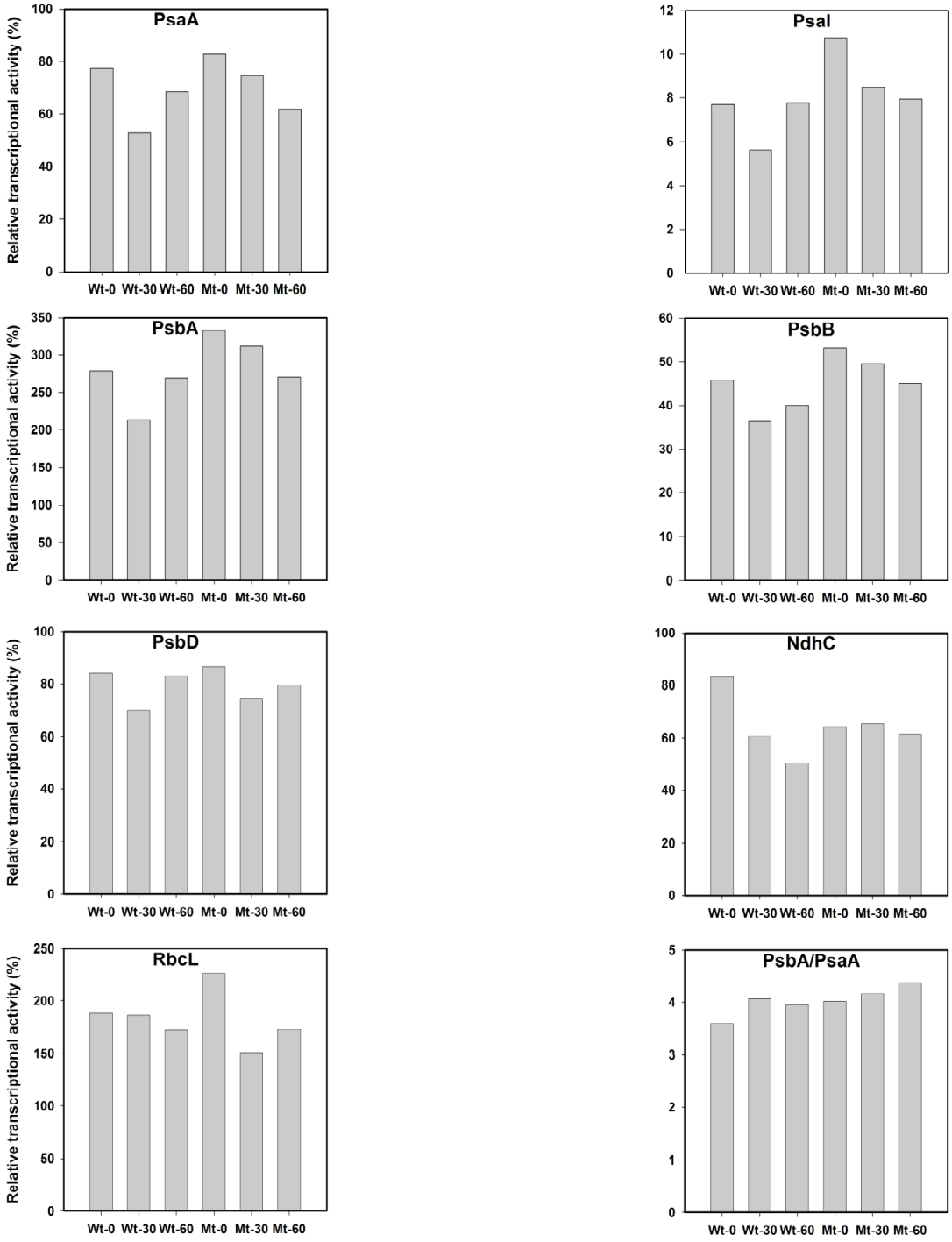


Figure 5.2 Transcriptional activities of selected chloroplast genes in wild type and CSK null mutant (SALK_027360) in response to short-term changes in light quality. Wt-0, Wt-30 and Wt-60 denote transcriptional activity in wild type at zero time (before changing light 1 with light 2) and at 30

minutes and at 60 minutes after changing light 1 with light 2. Likewise, Mt-0, Mt-30 and Mt-60 correspond to transcriptional activity in the CSK mutant at zero, 30 and 60 minutes of the light switch. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF, 2008).

mutant when compared to the wild type (Figure 5.3). Since most of the [α-^{32}P]UTP is incorporated in to the nascent mRNA species that were being transcribed in the run-on assay, it can be concluded that the total transcriptional activity of chloroplasts in response to the short-term light switch is higher in CSK mutants than in wild type.

In order to further characterize the transcriptional activity of chloroplast genes in CSK mutants, run-on assays were performed with four additional light conditions. This experiment is designed to assess the transcriptional activity in response to long-term changes in light quality. The light condition "Light 1" refers to growing plants continuously in light 1 for seven days; "Light 1 to 2", four days in light 1 before being transferred to light 2 for three days; "Light 2", seven days in light 2 and "Light 2 to 1", four days in light 2 and then transferred to light 1 for three days. Run-on assays for selected chloroplast genes were performed from samples taken at the end of the above light regimes. A few more chloroplast genes, additional to the ones chosen in the earlier experiment, were included in this assay. They were *ndhH*, encoding the H subunit of NAD(P)H dehydrogenase; *rpoB*, B subunit of the plastid encoded RNA polymerase (PEP); *clpP*, proteolytic subunit of protease Clp; *rps14*, S14 ribosomal protein and *trnE* encoding glutamyl-tRNA. The *psaI* gene, selected in the earlier assay, was not included in this analysis. The selection of *rpoB* and *clpP* genes was influenced by the observation that they are transcribed exclusively by the phage-type RNA polymerase (NEP) (Lysenko and Kuznetsov, 2005).

Results of the run-on assays revealed that, in the CSK mutants, the transcriptional activity of *psaA* and *psbA* genes were no different from that of wild type in any of the light conditions tested (Figure 5.4). The only notable difference in the CSK mutant, when compared to wild type, was in the transcriptional rate of *psbD* and

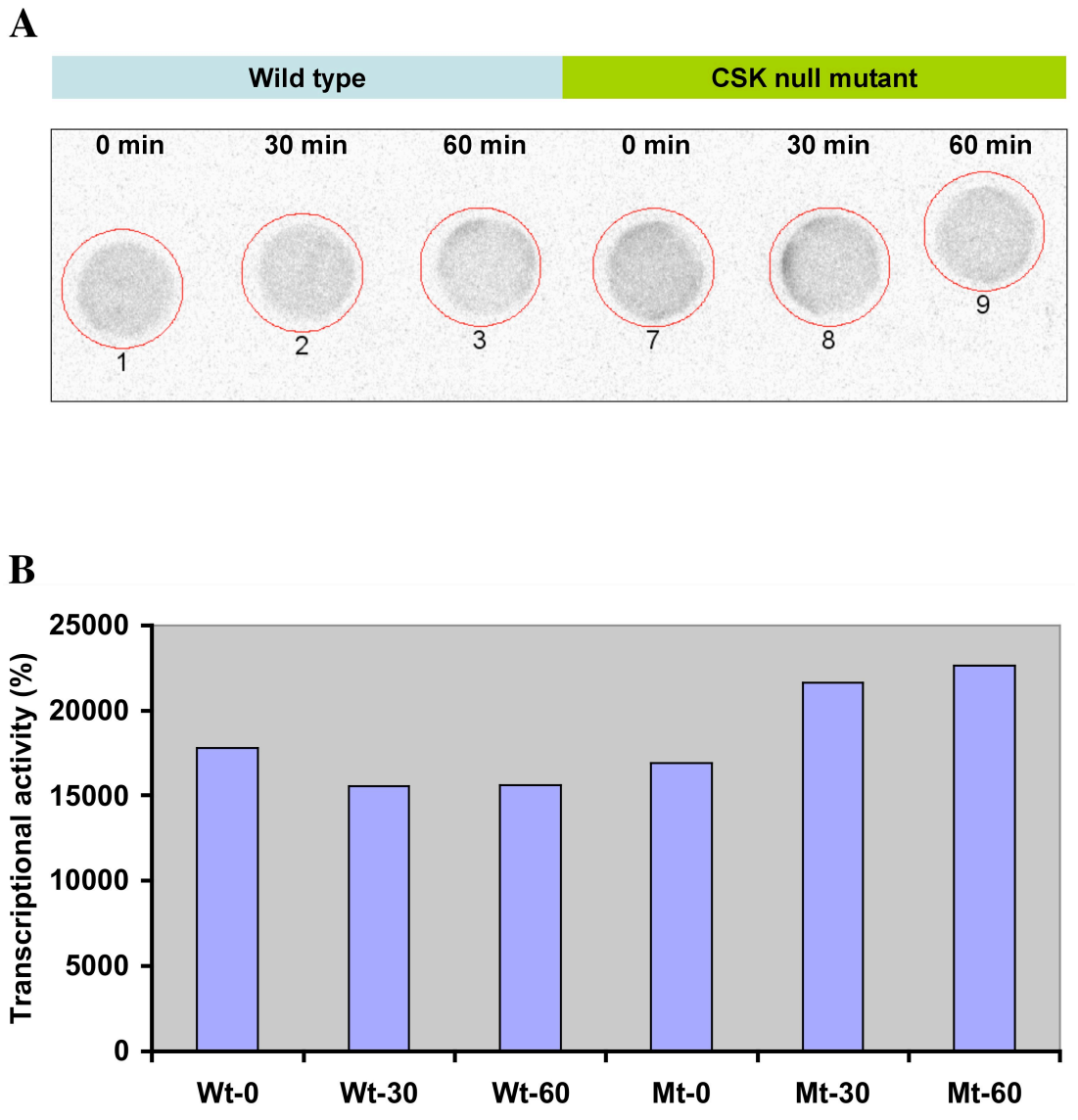


Figure 5.3 Total transcriptional activities of chloroplasts in wild type and CSK mutant (SALK_027360) in response to short-term changes in light quality. (**A**) The autoradiograph showing the signal intensities of the radio-labelled nascent chloroplast transcripts in wild type and CSK mutant at zero time (before replacing light 1 with light 2) and at 30 and 60 minutes after replacing light 1 with light 2. (**B**) Signal intensities plotted as a bar diagram. Total transcriptional activity of chloroplasts (in percentage) is shown on the Y-axis and samples on X-axis. Abbreviations used for samples, Wt-0, Wt-30, Wt-60, Mt-0, Mt-30 and Mt-60, are same as in Figure 5.2. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF, 2008).

ndhC genes (Figure 5.4 and figure 5.5). These genes showed an altered transcriptional activity in response to Light 1 to 2 and Light 2 to 1 switches in the CSK mutant (Figure 5.4 and figure 5.5). The transcript abundance of *ndhC* when quantified with northern blotting confirmed the altered transcriptional rate seen in the run-on assay (Figure 5.6).

The sum total of the transcriptional activity of all the chosen chloroplast genes showed that the overall chloroplast transcriptional activity was comparatively higher in the CSK mutant (Figure 5.7). This demonstration is consistent with the earlier observation of the elevated total chloroplast transcriptional activity of the CSK mutant in response to short-term light 1 to 2 switch (Figure 5.3). In order to eliminate the possibility that the higher transcriptional activity of the CSK mutant seen here resulted from the use of unequal amounts of chloroplasts in the run-on assay, equivalent amounts of chloroplast samples used for the run-on assay were checked on a SDS-PAGE gel (Figure 5.8). The SDS-PAGE of chloroplast samples suggested equal amounts of chloroplast proteins in all of the samples taken for the run-on assay (Figure 5.8). Thus the differences in transcriptional rates cannot be ascribed to unequal amounts of starting material in the assay.

5.4 Sequence features of CSK suggest mechanisms for redox sensing

As discussed in chapter 3, the sensor domain in CSK is likely to be formed by the ~180 amino acid long, N-terminally positioned GAF domain. The redox-sensing role of CSK, as suggested by the functional studies of CSK null mutants (Figure 5.1-5.8), would require a redox sensor domain in CSK. The N-terminal GAF domain could fulfill this function. A multiple sequence alignment of the GAF domain of CSK homologues revealed a positionally conserved cysteine residue (Figure 5.9). While GAF domains of CSKs in *Ostreococcus*, *Phaeodactylum* and a few *Synechococcus* species do not seem to have this cysteine conserved, other positionally conserved cysteines are present in their GAF domains (Figure 5.9). The multiple sequence alignment (Figure 5.9) also showed patches of conserved hydrophobic residues,

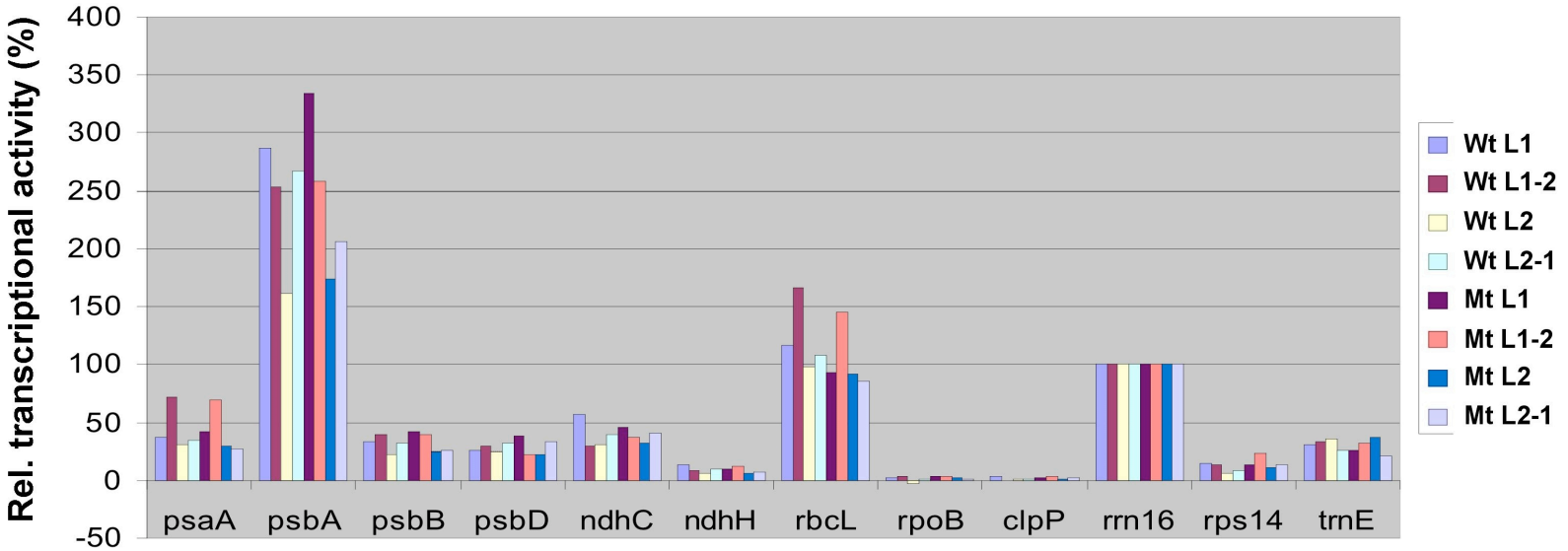


Figure 5.4 Transcriptional activities of selected chloroplast genes in wild type and CSK mutant (SALK_018074) in response to long-term changes in light quality. Following abbreviations, Wt L1, Wt L1-2, Wt L2, Wt L2-1, Mt L1, Mt L1-2, Mt L2 and Mt L2-1, are used for light conditions, Light 1, Light 1 to 2, Light 2 and Light 2 to 1, for wild type (Wt) and CSK mutants (Mt) respectively. Transcriptional activities of individual genes were normalized to that of *rrn16* gene. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF 2008).

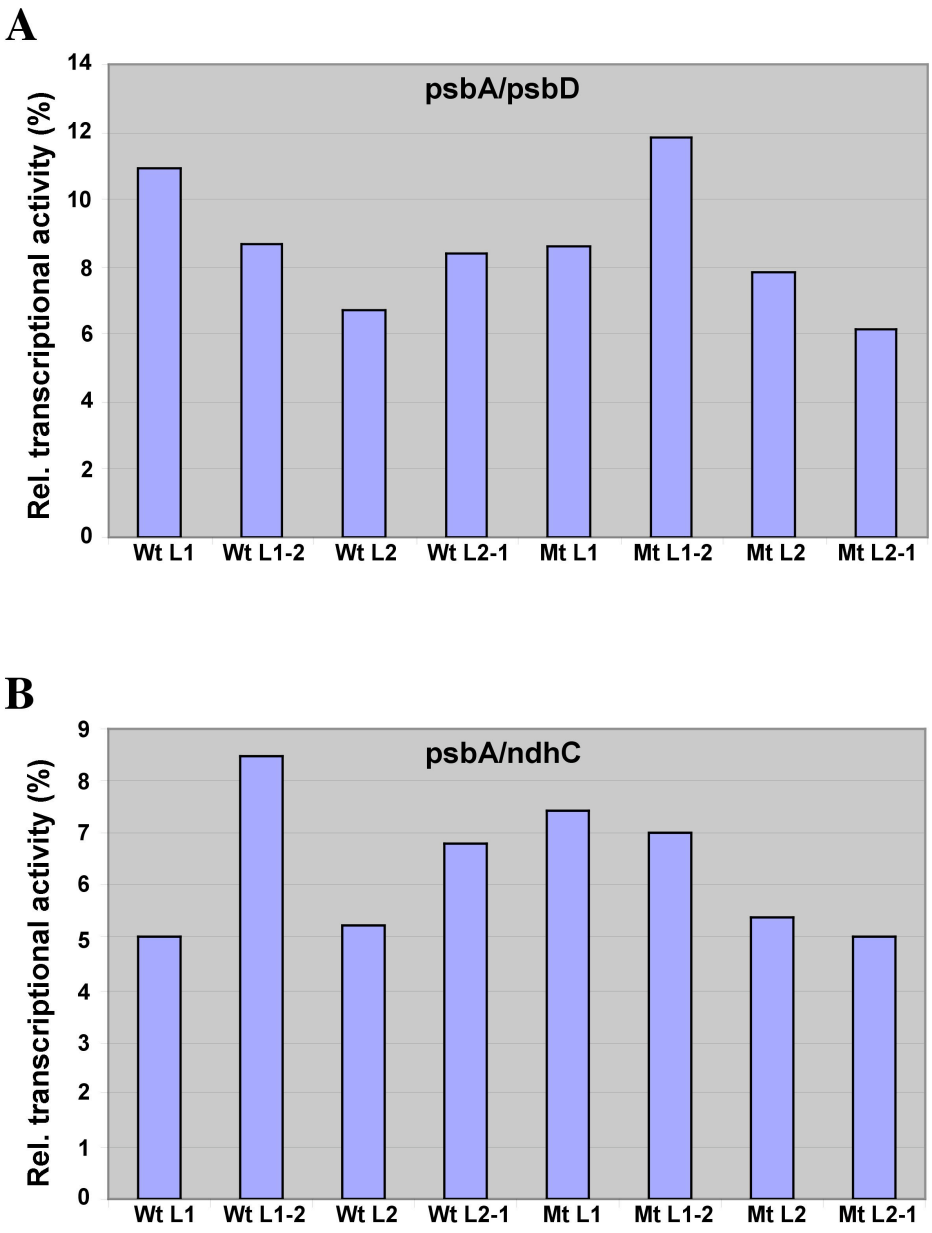


Figure 5.5 Transcriptional activities of *psbD* and *ndhC* genes are represented as ratios to the *psbA* gene. (**A**) Transcriptional activity of *psbD* gene represented as ratio to the *psbA* gene. (**B**) Transcriptional activity of *ndhC* gene represented as ratio to the *psbA* gene. Abbreviations for samples are same as in Figure 5.4. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF 2008).

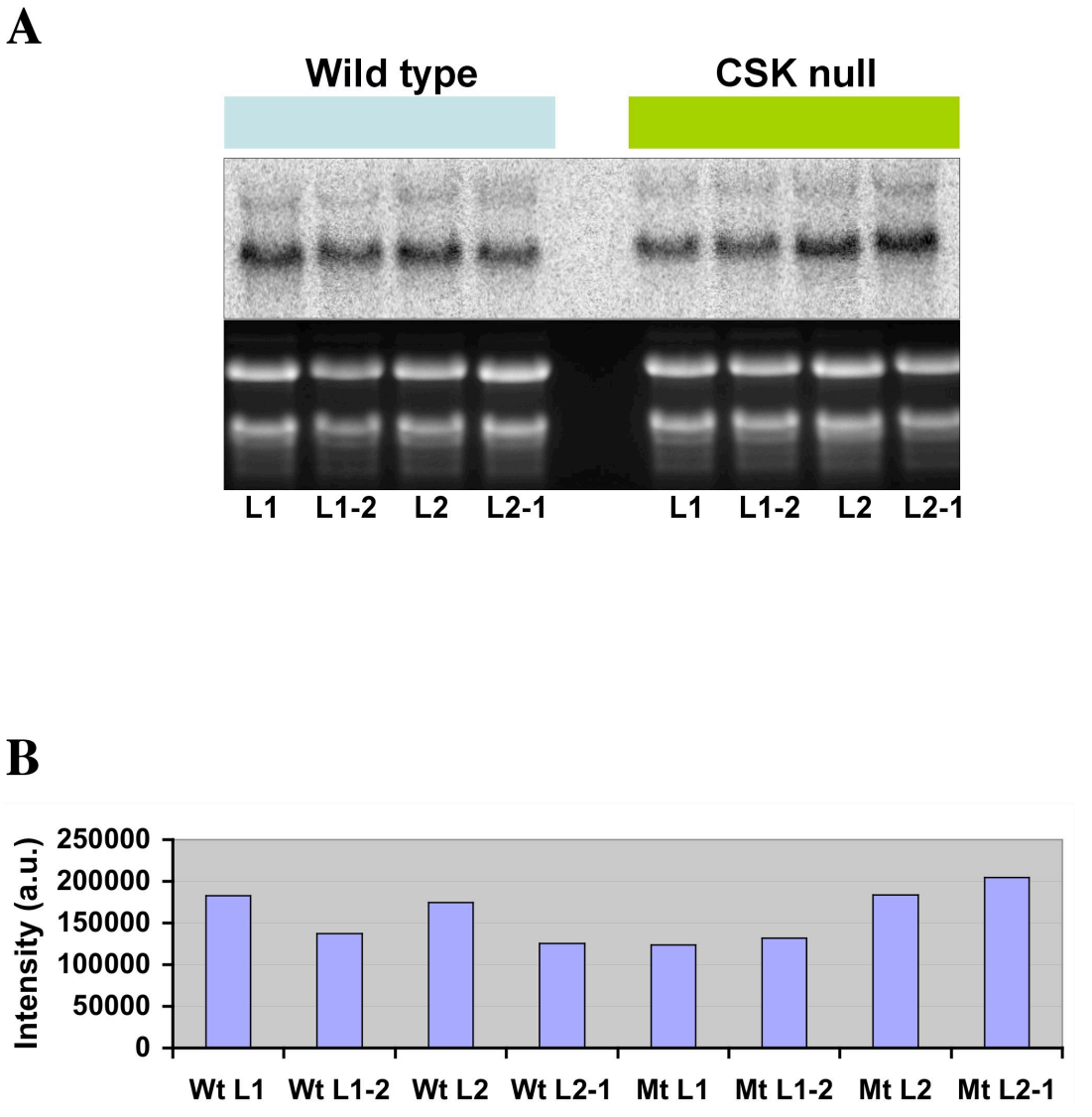


Figure 5.6 Transcript abundance of *ndhC* gene as quantified with northern blotting. (**A**) The autoradiograph showing signal intensities of the *ndhC* transcript in different light conditions. Lower panel shows equal loading of RNA (ribosomal) in each sample. Samples, L1, L1-2, L2 and L2-1, correspond to light conditions, Light 1, Light 1 to 2, Light 2 and Light 2 to 1, for wild type and CSK mutants. (**B**) Signal intensities of the ndhC transcript presented as a histogram. Following abbreviations, Wt L1, Wt L1-2, Wt L2, Wt L2-1, Mt L1, Mt L1-2, Mt L2 and Mt L2-1, are used for light conditions, Light 1, Light 1 to 2, Light 2 and Light 2 to 1, for wild type (Wt) and CSK mutants (Mt) respectively. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF 2008).

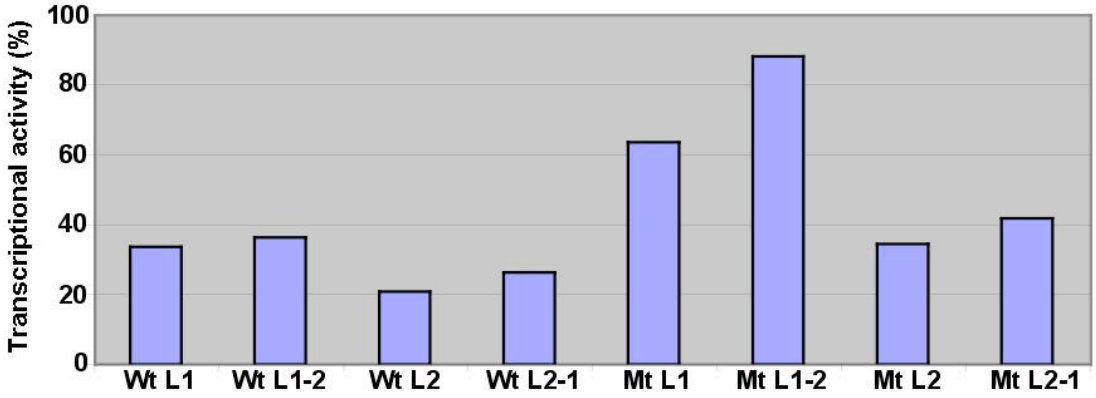


Figure 5.7 Sum total of transcriptional activities of all the selected chloroplast genes in wild type and CSK mutant (SALK_018074) in response to long-term changes in light quality. The values presented here were derived from the data shown in Figure 5.4. Total transcriptional activity for each sample is calculated by adding transcriptional activities of all the selected genes in the "long-term light switch experiment." Abbreviations used to denote samples are same as in Figure 5.6B. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF 2008).

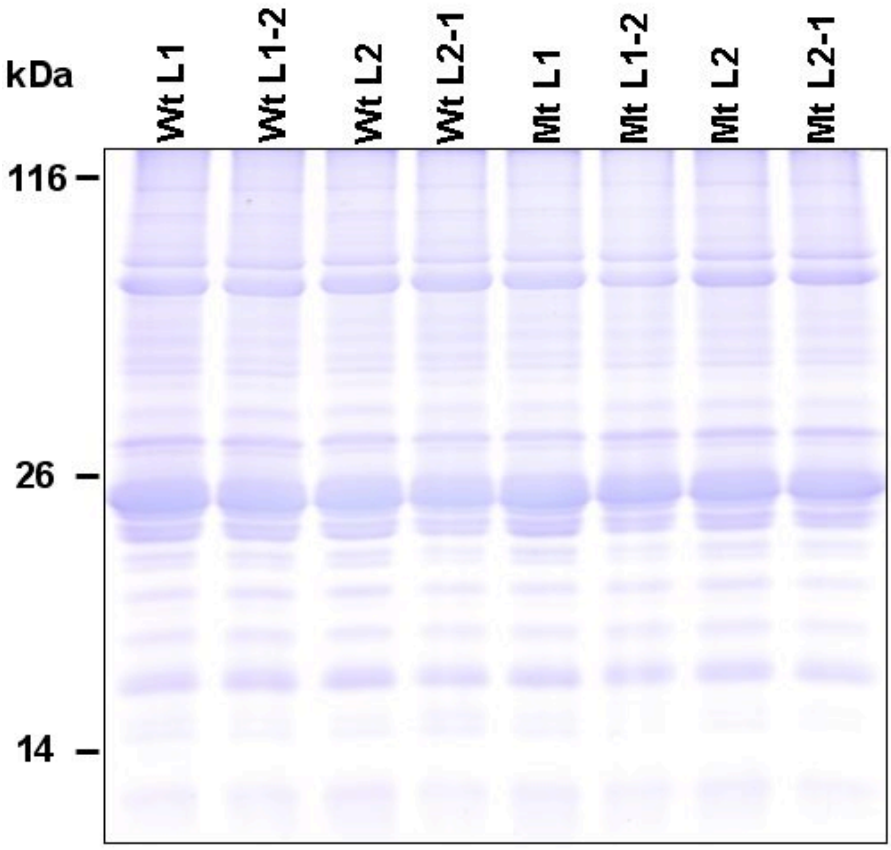


Figure 5.8 SDS-PAGE shows that chloroplasts used for wild type and CSK mutant samples in the "long-term" run-on assay were equal in amounts. Equivalent amounts of chloroplasts used for the run-on assay were checked on a 10% SDS-PAGE gel and stained with Coomassie. Molecular weight markers (in kDa) are indicated on the left. The following abbreviations, Wt and Mt, were used for wild type and CSK mutant respectively. L1, L1-2, L2 and L2-1 denote light conditions, Light 1, Light 1 to 2, Light 2 and Light 2 to 1 respectively. (Unpublished results, Puthiyaveetil S., Dietzel L., Pfannschmidt T. and Allen JF 2008).

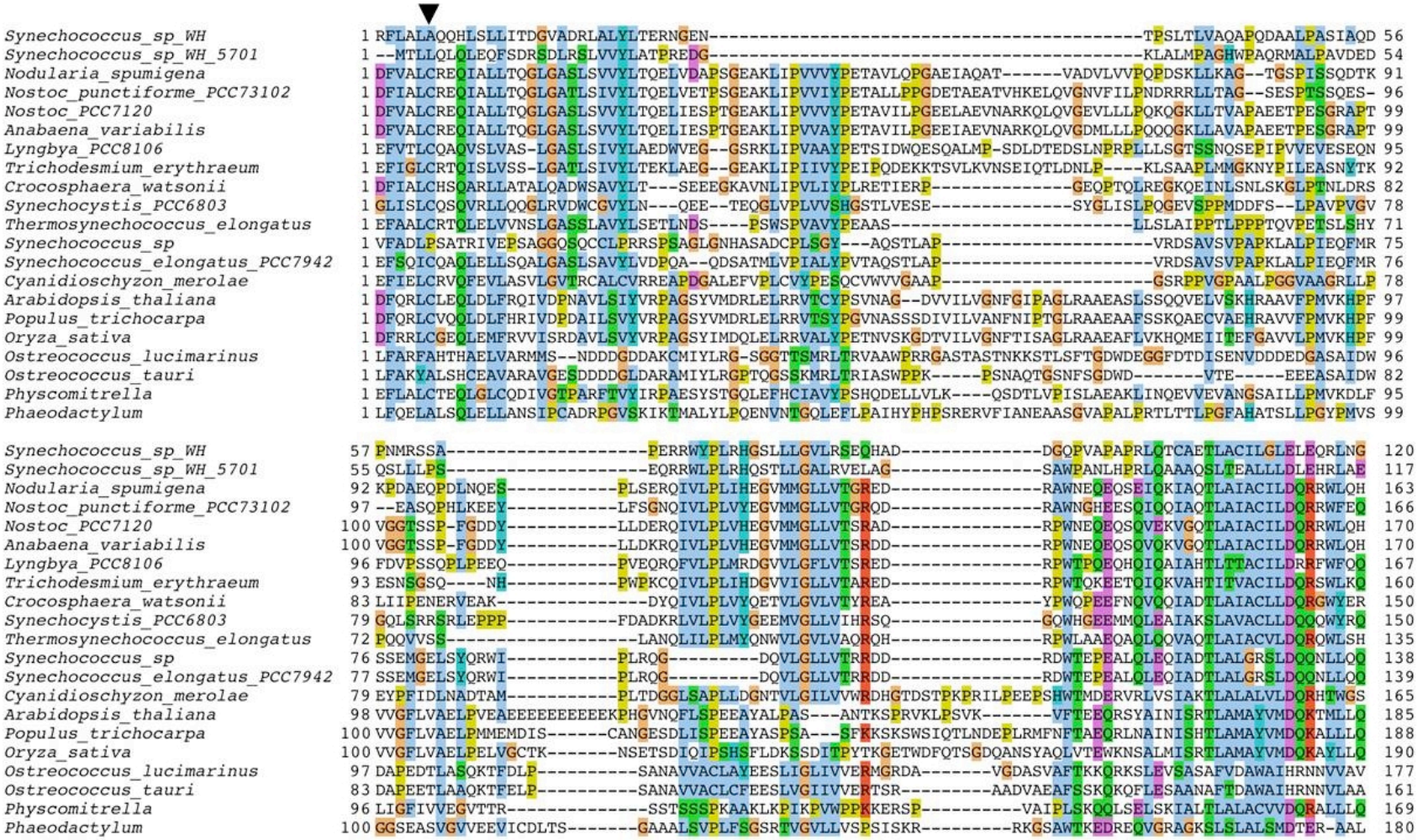


Figure 5.9 Multiple sequence alignment of the predicted GAF domain of the *Arabidopsis thaliana* (CSK) and its plant, algal and cyanobacterial homologues. The residue colouring scheme is same as in Figure 3.1. The conserved cysteine residue in the cyanobacterial and plant CSKs is indicated by the arrowhead.

which resemble aromatic prosthetic group binding sites in proteins, in the GAF domain of CSK. A sequence similarity search with the CSK GAF domain, using the PSI BLAST program (Altschul et al., 1997), revealed distant homology to the nucleotide binding sites of some metabolic enzymes.

5.5 Discussion

The loss of CSK alters the kinetics of light-quality driven transcript accumulation in chloroplasts both qualitatively and quantitatively (Figure 5.1). Since the light-quality induced transcriptional response is triggered by changes in the photosynthetic electron transport chain (Fey et al., 2005b, Pfannschmidt et al., 1999a, Puthiyaveetil and Allen, 2008, Tullberg et al., 2000), I propose that CSK couples photosynthetic electron transport with chloroplast gene transcription. The complex kinetics and transients of the chloroplast transcript accumulation (Figure 5.1) have been attributed to the superimposition of an endogenous rhythm of mRNA abundance to the light quality driven transcription (Puthiyaveetil and Allen, 2008). The endogenous rhythm of chloroplast transcript abundance is shown to be independent of circadian rhythm and is postulated to be dependent on the activity of the photosynthetic electron transport chain (Doran and Cattolico, 1997). Since both the light quality and oscillatory components of transcript accumulation are affected in CSK mutants, the two responses, light quality and oscillatory, may therefore be governed by the CSK-based mechanism (see chapter 9 for further discussion).

Light quality driven chloroplast transcriptional response is part of a reprogramming of the thylakoid composition known as photosystem stoichiometry adjustment, as discussed in chapter 1. In *Chlamydomonas reinhardtii*, adjustment of photosystem stoichiometry is shown to be achieved through suppression of the biosynthesis of PS I components (Murakami et al., 1997a), similar to that demonstrated for the cyanobacterium, *Synechocystis* PCC 6714 (Fujita, 1997). If this is how the photosystem stoichiometry adjustment is accomplished in photosynthetic organisms, the absence of *psaA* transcriptional repression in the early transients of

the light 2 to 1 switch in CSK mutants (Figure 5.1B) strengthens the suggested role of CSK in photosystem stoichiometry adjustment. The aberrant chlorophyll *a*/*b* ratios in CSK mutants in response to light quality changes (Table 5.1) are further indicative of CSK's role in regulating photosystem stoichiometry. It should be mentioned that the transcript accumulation kinetics of the *psbA* gene, encoding the D1 reaction centre apoprotein of PS II, were found to be too subtle and far less pronounced than those of the *psaA* gene in response to light switches (Puthiyaveetil and Allen, 2008). Hence transcript accumulation kinetics of the *psbA* gene is not included in the analysis of the transcript accumulation kinetics reported here.

The preliminary data on chloroplast run-on transcription analysis (Figure 5.2) demonstrates that the CSK mediated gene regulation in chloroplasts operates through differential promoter usage. The target genes of the CSK-based transcriptional activation/repression mechanism involve photosystem genes as well as non-photosystem genes (Figure 5.2). Although the exact number of genes under the control of the CSK regulatory switch remains to be determined, there are clear indications from the run-on analysis (Figure 5.2) that the switch controls a set of genes rather than one or two target genes. This regulatory feature of CSK is in tune with the demonstrated properties of two-component systems, which usually exert control over a few distinct sets of target genes by both transcriptional activation and repression mechanisms (Li et al., 2008, Lynch and Lin, 1996).

Although run-on analysis of the transcriptional activation/repression of chloroplast genes in response to short-term changes in light quality (Figure 5.2) helped gain some insights into the operation of CSK, the kinetics of transcriptional activity for the wild type controls reported here (Figure 5.2) differed from the earlier report on mustard (Pfannschmidt et al., 1999a). In mustard, in response to light 1 to 2 switch, the *psaA* transcriptional activity increased up to four fold, from that present in light 1 adapted plants, in half an hour from the light switch (Pfannschmidt et al., 1999a). On the same time scale, the wild type control described here showed an initial slump in *psaA* activity from that present before the light switch, which then

seemed to be rising (Figure 5.2). The *psbA* transcriptional activity, likewise, differed from than seen for mustard (Pfannschmidt et al., 1999a).

It is not clear whether this apparent difference between *Arabidopsis thaliana* and *Sinapis alba* represents a true species-specific difference in acclimatory response. It has been known the chloroplast transcriptional activity declines with the age of the plant (Mullet and Klein, 1987). *Arabidopsis* plants used for the run-on assay here were at least 5 weeks old, while the mustard plants used in the earlier report (Pfannschmidt et al., 1999a) were only 7 days old. When using *Arabidopsis*, it is difficult to manage the trade-off between age of the plant and the amount of leaf material needed for the run-on assay. The slower transcriptional response seen in *Arabidopsis* may thus be attributed to the greater age of the experimental material. However, a similar, slow transcriptional response to changes in light quality has been observed in pea chloroplasts that were isolated from 8 days old plants (Tullberg et al., 2000). It is also interesting to note that the slower *Arabidopsis* transcriptional response seen in the run-on analysis is consistent with the early transients of transcript accumulation reported with the qPCR technique (Figure 5.1) (Puthiyaveetil and Allen, 2008). Additionally, in the qPCR analysis, *Arabidopsis* plants were only 2 weeks old. Nevertheless more biological and technical replicates should be used before solid conclusions can be drawn from the run-on analysis reported here. Chloroplast run-on analysis should also be performed for samples from light 2 to 1 switch. Additionally, if feasible, run-on assays should be performed with young Arabidopsis plants (2-3 weeks old) to rule out age-related effects on chloroplast transcriptional activity.

Run-on analysis of the long-term light switch samples showed some unexpected, but interesting regulatory features of CSK (Figure 5.4). The wild type-like transcriptional activities of *psaA* and *psbA* genes in CSK mutants were a surprising observation (Figure 5.4). However, it should be recalled that the "long-term" run-on assays were performed after seven continuous days of light acclimation for Light 1 and Light 2 conditions and three continuous days for the switched Light 1

to 2 and Light 2 to 1 conditions. The run-on assays for determining the kinetics of transcriptional activity in response to a short-term light switch, in contrast, were performed on a time-scale of minutes. The long-term acclimation extending to days will surely involve multiple factors. These observations, however, support the possibility that the CSK-based regulatory mechanism governs the rapid induction of transcription rather than its long-term maintenance over the time-scale associated with photosystem stoichiometry adjustments (Schutze et al., 2008), and that additional components may be involved in regulating photosystem stoichiometry. The timescale of the CSK mediated chloroplast transcriptional response suggested here is also in agreement with the timescale of transcriptional regulation that is characteristic of two-component systems in bacteria.

The consequence of the altered *psbD* and *ndhC* transcriptional activity in CSK mutants under long-term changes in light quality is unclear (Figure 5.4 and figure 5.5). The regulation of *psbD* gene, which encodes for the D2 subunit of PS II, can obviously be the part of a photosystem stoichiometry adjustment. Transcriptional regulation of *psbD* has also been reported in other physiological contexts. It has been known that a unique blue-light responsive promoter mediates light-dependent transcription of *psbD* gene and that in *Arabidopsis* (Kim et al., 1999), the nuclear-encoded sigma factor, AtSig5 is essential for the recognition of the blue-light dependent promoter by the plastid encoded polymerase (PEP) (Tsunoyama et al., 2004). The *PsbD* gene is also under the control of an endogenous circadian oscillator (Nakahira et al., 1998). Chloroplast NAD(P)H dehydrogenase, whose subunit D3 is encoded by the *ndhC* gene, has been implicated in the chlororespiratory pathway (Peltier and Cournac, 2002). It is believed that the PQ pool, in addition to transferring electrons from PS II to cytochrome b6f, also acts as a conduit for electrons from the NAD(P)H dehydrogenase to cytochrome b6f complex as part of a suggested chlororespiratory pathway (Peltier and Cournac, 2002). It then appears that under conditions of preferential illumination of PS II, the genes encoding the components of NAD(P)H dehydrogenase may be downregulated, since their products would otherwise add electrons to an already reduced quinone pool. This possibility,

however, cannot account for all the data in Figure 5.6. Further research is required in order to address the functional significance of *ndh* gene regulation in chloroplasts.

An important observation made from the run-on analysis of both the "short-term" and "long-term" experiments is that the inactivation of *CSK* gene causes an increase in the overall transcriptional activity of chloroplasts (Figure 5.3 and figure 5.7). While this may have arisen from the loss of regulation of some chloroplast genes in the absence of a functional CSK, the exact nature of the causes and consequences of the elevated chloroplast transcriptional activity in CSK mutants remain to be understood.

The functional characterisation of CSK knockout mutants (Figure 5.1-5.8) clearly suggests that CSK, despite being a modified histidine kinase in green algae and plants, continues to function as a sensor component of a two-component regulatory system in chloroplasts. The transcriptional regulatory role of CSK in red algae and diatoms, species in which CSK exists as an unmodified histidine kinase, is attested by the presence of chloroplast response regulators, which are DNA-binding transcription factors (Duplessis et al., 2007). This then brings us to the following question: what mediates the transcriptional regulatory role of CSK in green algae and plants? The identity and the action of the response regulator partner of green algal and plant CSKs are discussed in chapter 8.

Since the light quality driven chloroplast transcriptional response is initiated by changes in the redox state of an electron carrier located between two photosystems (Pfannschmidt et al., 1999a, Puthiyaveetil and Allen, 2008, Tullberg et al., 2000) and a functional CSK is required for this process (Figure 5.1-5.8), it appears that CSK acts as a redox sensor in chloroplasts. Additionally, the conserved GAF sensor domain in CSK displays features of a redox sensor domain (Figure 5.9). A GAF domain forms the redox sensor domain in the DosS redox sensor of *Mycobacterium tuberculosis* (Kumar et al., 2007). A haem prosthetic group confers the redox sensing ability of the DosS GAF domain. A conserved redox-responsive

cysteine residue in the GAF domain of a bacterial phytochrome has aided its functional transformation from light sensor (photoreceptor) to redox sensor (Vuillet et al., 2007). GAF domains also show structural homology and functional overlap with the well-characterized redox sensor input domain, PAS (Ho et al., 2000, Sardiwal et al., 2005).

The conserved cysteine residue in the GAF domain of CSK may act as a redox signal reception site (Paget and Buttner, 2003), but the lack of its complete conservation among CSK homologues weakens this possibility (Figure 5.9). The conserved hydrophobic patches of residues in the GAF domain of CSK (Figure 5.9) together with its remote homology to the nucleotide binding sites of some enzymes raise the possibility that the GAF domain in CSK employs redox-responsive prosthetic groups such as flavin nucleotides for its redox-sensing function. The presumed redox sensing functional role of CSK is further supported by the recent demonstration that the cyanobacterial homologue of CSK (discussed in chapter 7) interacts with a redox response regulator in cyanobacteria called rppA (Sato et al., 2007). It also emerged that the cyanobacterial homologue of CSK, Hik2, interacts with a component of the photosynthetic electron transport chain, the phycobilisome linker protein, apcE (Sato et al., 2007).

This then takes us to the nature of the redox signal itself. The identity of the redox signal that switches CSK on or off is yet to be revealed. Since the light quality driven, CSK-mediated transcriptional response is initiated by changes in the redox state of plastoquinone (Pfannschmidt et al., 1999a, Puthiyaveetil and Allen, 2008, Tullberg et al., 2000), it is reasonable to assume that the signal perceived by CSK is the redox state of PQ itself. The signalling property of quinone electron carrier is well documented in literature (Allen et al., 1981, Georgellis et al., 2001, Grabbe and Schmitz, 2003, Pfannschmidt et al., 1999a, Ivleva et al., 2006, Escoubas et al., 1995, Swem et al., 2006, Hihara et al., 2002). This, however, raises the question of how the redox state of an electron carrier such as plastoquinone, within the thylakoid membrane, might be sensed by an apparently soluble redox sensor such as CSK?

Sensing redox signals originating from the bioenergetic membrane seems to be difficult for soluble redox sensors. But examples in bacteria suggest that it is possible. The light input component of cyanobacterial circadian oscillator, CikA, is a soluble protein sensing PQ redox state (Ivleva et al., 2006). The NifL sensor, which regulates the expression of nitrogen fixing genes in *Azotobacter vinelandii*, is another example of a soluble redox sensor of quinone (Grabbe and Schmitz, 2003). A third example is the antirepressor protein AppA, which regulates photosynthesis gene regulation in *Rhodobacter sphaeroides* (Oh and Kaplan, 2000). Although PQ seems to be an interesting candidate for the signal affecting CSK activity, stromal electron carriers such NADP, thioredoxin and glutathione cannot be excluded at this stage. Further studies are indeed required to pinpoint the exact identity of the signal. The use of specific electron transport inhibitors will be a useful direction in this investigation.

6.1 Introduction

Signal perception and concomitant auto-phosphorylation of sensor kinases are initial steps in the detection and transduction of signals by two-component signalling pathways (Stock et al., 2000). Sensor kinases autophosphorylate at a conserved histidine residue in the catalytic core. This histidine autophosphorylation site in CSK, barring exceptions in the red alga, *C. merolae* and diatoms, *P. tricornutum* and *T. pseudonana*, has been replaced with a tyrosine residue in the green alga, *Ostreococcus* and with a glutamate residue in all lower and higher plants (chapter 3). Red algal and diatom CSKs are thus likely to function as bacterial-type sensor histidine kinases as they continue to exist as unmodified sensor histidine kinases. In contrast, the green algal and plant CSKs have clearly undergone molecular evolution in their kinase domain and are now modified sensor histidine kinases.

Similar molecular evolution in the kinase domain has been reported for other sensor kinases. The DivL sensor, which is involved in cell division and differentiation in *Caulobacter crescentus*, has a tyrosine residue replacing the conserved histidine autophosphorylation site (Wu et al., 1999). This replacement of histidine with tyrosine has resulted in tyrosine kinase activity in the DivL sensor. Plant phytochromes and ethylene sensors are also modified histidine kinases with their conserved histidine residues replaced by other amino acids. Plant phytochromes and ethylene sensors are now known to function as serine/threonine kinases (Moussatche and Klee, 2004, Yeh and Lagarias, 1998). The altered phosphoryl group chemistry in DivL, plant phytochromes and ethylene sensors has not affected their signalling role as they continue to regulate various physiological responses. The signalling and functional properties of CSK have likewise not been compromised (chapter 5) despite CSK having undergone molecular changes in its kinase domain. Additionally, the highly conserved nature of the ATP binding domain in CSK (chapter 1) argues for its continued existence as a protein kinase. This possibility is now examined in the *Arabidopsis* CSK. The experimental results obtained and the inferences drawn are discussed.

6.2 Overexpression and purification of CSK-GST fusion protein in *E. coli*

Heterologous expression and purification in bacterial systems have been widely used for biochemical and structural studies of proteins. While bacterial expression systems work well for many proteins, it is challenging to express certain foreign proteins, such as protein kinases, in bacteria. The stability of the recombinant plasmid, toxicity, and solubility of foreign proteins are some of the problems encountered in bacterial systems. Nevertheless, the relative ease with which bacteria can be genetically manipulated, combined with their short doubling time, make them an ideal system for expression and purification of foreign proteins.

Autophosphorylation assays for histidine kinases are usually performed with their overexpressed, purified kinase domain. The kinase domain of sensor histidine kinases exhibits constitutive histidine kinase activity (autophosphorylation) in the absence of their sensor domains. This arises from the fact that, when the relevant signal or stimulus is absent, the sensor domains impose an inhibitory effect on the activity of the kinase domain. Therefore, sensor domains are usually not included in the protein construct used for the autophosphorylation assay. In order to study the enzyme activities of CSK *in vitro*, a partial cDNA clone that encodes for the kinase domain of CSK (residues, 150-611) was cloned into a GST-based expression system (Figure 6.1). This protein construct includes, in addition to the complete kinase domain, 155 amino acid residues N-terminal to the kinase domain of *Arabidopsis* CSK (part of the GAF sensor domain). This was to include any putative autophosphorylation site lying outside the HisKA domain of *Arabidopsis* CSK as it had lost its conserved histidine autophosphorylation site.

This truncated version of CSK was successfully overexpressed in *E. coli* as a GST fusion protein (Figure 6.2). Majority of the overexpressed CSK-GST fusion protein accumulated in the insoluble fraction, nevertheless, some CSK-GST proteins were still found in the soluble fraction (Figure 6.3). The CSK-GST fusion protein in

the soluble fraction was purified by affinity chromatography (Figure 6.3). Additives that prevent aggregation were used in all buffers and solutions in order to preserve the stability and activity of CSK-GST fusion proteins (chapter 2, Materials and Methods).

6.3 CSK autophosphorylates *in vitro*

In order to test whether CSK becomes auto-phosphorylated *in vitro*, the overexpressed and purified CSK-GST fusion was incubated in the presence of [γ-^{32}P]ATP, the reducing agent dithiothreitol, and the divalent cations Mg^{2+}, Mn^{2+} and Ca^{2+} (Chapter 2, Materials and Methods). It is known that bacterial sensor kinases have specific requirements for certain divalent cations for autophosphorylation and signal perception (Stock et al., 2000). The CSK protein becomes autophosphorylated in the presence of Mn^{2+} and to a lesser degree in the presence of Ca^{2+}, but no autophosphorylation was detected in the presence of Mg^{2+} (Figure 6.4A). Dithiothreitol stimulates autophosphorylation but is not an absolute requirement. Figure 6.4B shows little or no labelling of the GST protein alone by [γ-^{32}P]ATP under any of these conditions, confirming that the labelling results from autophosphorylation of CSK itself. A monoclonal antibody directed against the GST tag confirmed the identity of the ^{32}P-labeled band as of CSK-GST fusion protein (results not shown).

6.4 Alkaline phosphatase treatment confirms that specific labelling with ^{32}P results from phosphorylation of one or more amino acid side chains of CSK

To further confirm that the labelling of CSK results from specific phosphoryation of an amino acid side chain by ^{32}P, the labeled CSK proteins were incubated with calf intestine alkaline phosphatase, which hydrolyzes orthophosphoric acid monoester groups (Coleman, 1992). Alkaline phosphatase treatment completely removed labelling of CSK (Figure 6.4C). This confirms that the CSK is labeled with ^{32}P as a

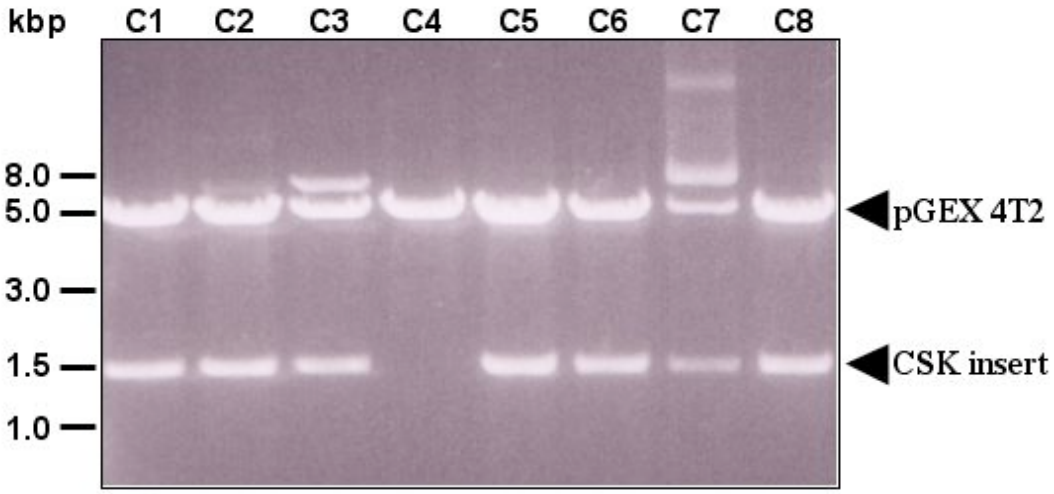


Figure 6.1 Screening for recombinant pGEX 4T2 plasmids containing the CSK insert. Recombinant clones (C1 to C8) were double digested with EcoRI and NotI, reactions products separated on a 1% agarose gel and stained with ethydium bromide. Positions of bands from the DNA ladder (in kilo base pairs, kbp) are indicated on the left. All clones, except clone 4, seem to contain the CSK insert. Clones 2, 3 and 7 seem to indicate incomplete digestion.

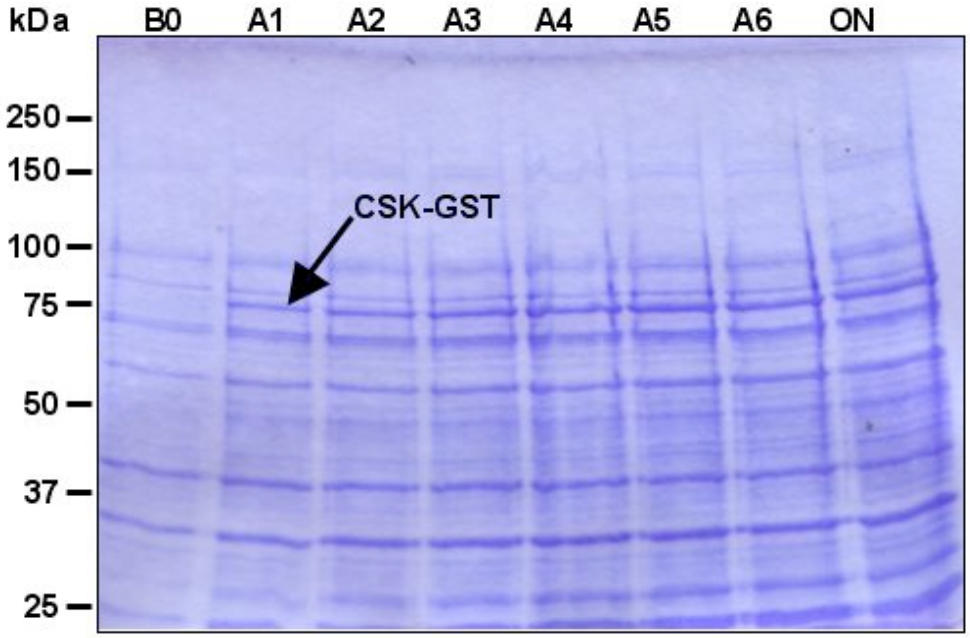


Figure 6.2 Overexpression of CSK-GST fusion protein in *E. coli*. A truncated version of CSK (462 amino acids, residues from 150 to 611) is overexpressed as a GST fusion protein in *E. coli*. The molecular weight of the truncated CSK is 50 kDa and of GST is 26 kDa. The expression of CSK-GST fusion protein was induced with IPTG. Bacterial cells were harvested before induction (B0) and after 1, 2, 3, 4, 5 and 6 hours after induction (A1-A6). "ON" corresponds to overnight from the induction. Bacterial cells were lysed and bacterial proteins were separated on a 12% SDS-PAGE gel and stained with Coomassie blue. Molecular weight markers are indicated on the left.

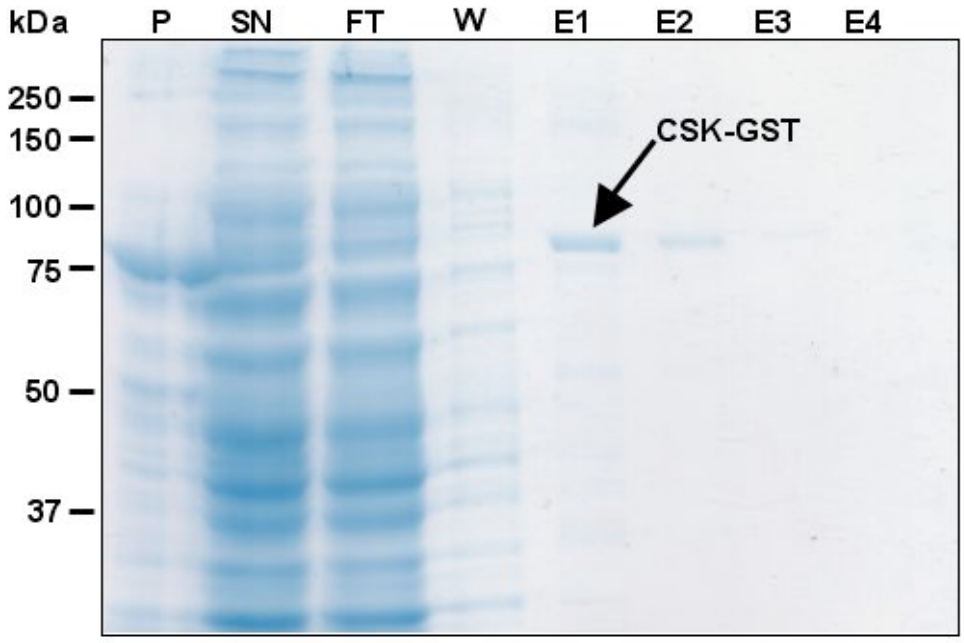


Figure 6.3 CSK-GST fusion protein at various stages of purification. Aliquots from various stages of purification were separated on a 12% SDS-PAGE gel and Coomassie stained. The following fractions were analysed on the gel: P, pellet; SN, supernatant; FT, flow through; W, wash; E1, first elution; E2, second elution; E3, third elution; E4, fourth elution. Positions of molecular weight markers are indicated on the left.

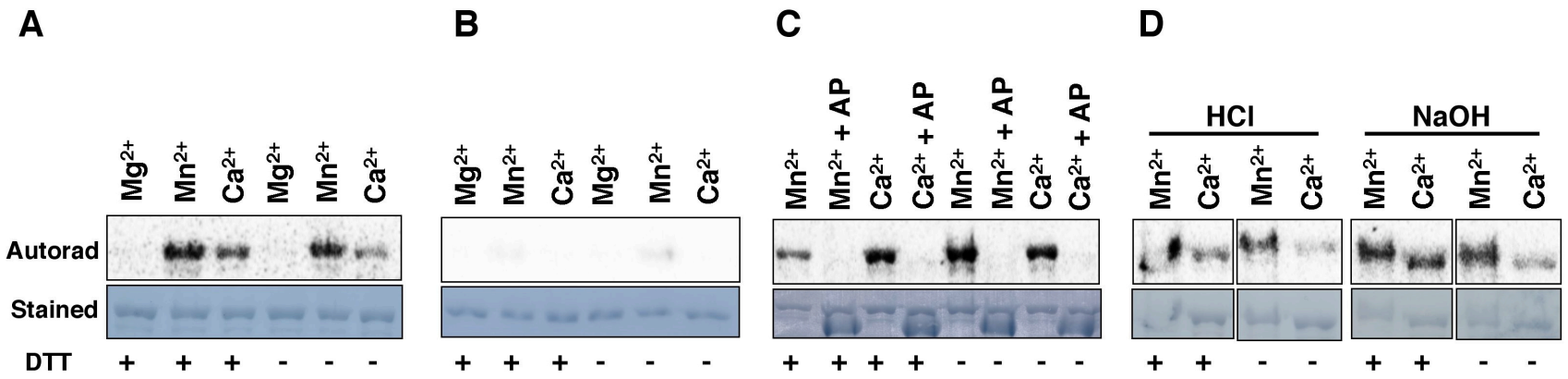


Figure 6.4 CSK is a phosphoprotein and appears to be its own protein kinase. (**A**) Over-expressed and purified CSK-GST fusion protein becomes autophosphorylated in the presence of Mn^{2+} or Ca^{2+} at 10 mM final concentration (**B**) GST protein alone is not labelled sufficiently to account for the labelling of CSK-GST. (**C**) Alkaline phosphatase treatment completely removes the labelling of the autophosphorylated CSK. (**D**) Acid/base stability assay shows CSK phosphoryl group is resistant to both acid and alkali treatments. "Autorad", the autoradiograph of the labelling reaction separated by SDS-PAGE. "Stained", the amido black staining of the corresponding PVDF membrane from which the autoradiograph was developed. The name of the divalent cation (Mg^{2+}/Mn^{2+}/Ca^{2+}) present in the labelling reaction is shown above each lane. "+AP" in panel C indicates treatment with alkaline phosphatase. The presence or absence of dithiothreitol, DTT, in the reaction medium at a final concentration of 2 mM is shown as + or - at the bottom of each lane.

result of autophosphorylation on an amino acid chain and not because of non-specific labelling such as non-covalent binding of ATP or covalent adenylation.

6.5 Acid and alkali-stability of phosphorylated amino acid(s) in autophosphorylated CSK suggests tyrosine phosphorylation

To determine the identity of the phosphorylated amino acid(s) in autophosphorylated CSK, the autophosphorylated protein was incubated in acid or alkali (Chapter 2, Materials and Methods). The stability or liability of phospho-amino group to acid or alkali treatment is a direct indicator of the nature of the phosphorylated amino acids in phosphoproteins. It is known that phosphohistidine is labile in acid but stable in alkaline solution, that phosphoserine and phosphothreonine are acid-stable and alkali-labile, and that phosphotyrosine is stable in both acid and alkali. The stability of the CSK phosphoryl group to both acid and alkali treatments (Figure 6.4D) suggest tyrosine phosphorylation.

6.6 Discussion

The specific labelling by ^{32}P in the autophosphorylation assay (Figure 6.4) clearly suggests that CSK autophosphorylates *in vitro*. The autophosphorylation of CSK is dependent on the presence of the divalent cation Mn^{2+} and to a lesser degree Ca^{2+} in the reaction medium (Figure 6.4A). While the bacterial sensor histidine kinases generally prefer the divalent cation Mg^{2+} for autophosphorylation reaction, Mn^{2+} and Ca^{2+} specific phosphorylation has been reported in some plant ethylene sensors, which are derived from sensor histidine kinases (Moussatche and Klee, 2004, Zhang et al., 2004).

The acid and alkali stability of the phosphorylated amino acid suggests autophosphorylation on a tyrosine residue in CSK (Figure 6.4D). Autophosphorylation on the conserved glutamate residue in the H-box or of a histidine residue lying outside the H-box is unlikely. This is because of the complete

dephosphorylation of phosphorylated CSK when treated with alkaline phopshatase (Figure 6.4C). Alkaline phosphatase treatment excludes phosphorylation on histidine or glutamate residues, as it is specific for hydrolysing *o*-phosphorylation (Coleman, 1992). Usually, the amino acids tyrosine, serine, and threonine undergo *o*-phosphorylation as a post-translational modification. Mn^{2+} specificity of CSK autophosphorylation also supports tyrosine kinase activity, as Mn^{2+} specificity is characteristic of tyrosine kinase autophosphorylation (Schinkmann and Blenis, 1997). Sequence features of green algal CSKs further support tyrosine phosphorylation as a tyrosine residue replaces their conserved histidine autophosphorylation site (chapter 3). Similar tyrosine replacement in DivL sensor results in tyrosine kinase activity (Wu et al., 1999). In plant CSKs, since a glutamate residue replaces the histidine (chapter 3), conserved tyrosine residues lying outside the H-box could be the site(s) of autophosphorylation of plant CSKs. Conserved tyrosine residues that could act as putative phosphorylation site(s) can be recognized in a multiple protein sequence alignment of plant CSKs.

Acid-hydrolysis of phosphorylated amino acids, followed by two-dimensional thin-layer electrophoresis can be further used to uncover the identity of phosphorylated amino acid(s) in CSK. Mass spectrometry-based approaches can also be employed to map the phosphorylation site(s) in the autophosphorylated CSK. MS analysis of proteolytic fragments of autophosphorylated CSK coupled with microsequencing will reveal the identity of phospho amino acid(s) in CSK.

The CSK protein used in the autophosphorylation assay here is a truncated version, which does not contain the first 149 amino acids from the N-terminus of the protein. The missing N-terminal segment also contains the serine rich transit peptide, which is believed to be part of the mature protein as suggested by the chloroplast protein import assays (chapter 4). The putative function of the retained transit peptide in CSK as an autophosphorylation site (discussed in chapter 4) should be tested by autophosphorylation assay with the full-length version of the CSK protein. The truncated version of CSK does not contain the complete GAF sensor domain,

which is likely to form the redox-sensor input domain in CSK (discussed in chapter 5). Although the reducing agent, DTT, is found to stimulate autophosphorylation in the truncated version of CSK used in our study, the redox dependency of autophosphorylation should be examined in the full-length version of CSK.

It should be noted that, even though the ^{32}P labeling observed in my experimental condition is best explained by the autophosphorylation activity of CSK, phosphorylation of CSK by a copurifying bacterial kinase (from *E. coli*) cannot be completely excluded at this stage. However, this seems unlikely, as CSK require Mn^{2+} instead of the divalent cation, Mg^{2+}, which is usually preferred by bacterial sensor kinases. Autophosphorylation assays with CSK protein constructs carrying point mutations in the conserved motifs (N, G1, F and G2) of the ATP binding domain should be performed to further test the nature of the ^{32}P labeling of CSK.

It remains to be seen whether the demonstrated autophosphorylation activity of CSK also occurs *in vivo* and whether autophosphorylation is required for CSK function. But at least one thing is clear; CSK, despite undergoing molecular evolution in its kinase domain, continues to function as an archetypical bacterial sensor kinase and that it employs phosphoryl group transfer as its signalling mechanism *in vitro*.

7.1 Introduction

Photosynthetic eukaryotes consist of a vast array of organisms as different from each other as a free floating unicellular alga and an oak tree. Nevertheless, they have one thing in common; their plastids. It is believed that plastids of all photosynthetic eukaryotes originated, at around 1.5 billion years ago, from a singular symbiotic event involving an eukaryotic host cell and a cyanobacterium (Gould et al., 2008). The monophyly and the cyanobacterial ancestry of chloroplasts have direct bearing on the phylogeny and distribution of CSK in photosynthetic eukaryotes. If the prediction is correct that CSK is a genetic control mechanism inherited from the bacterial ancestor of chloroplasts (Allen, 1993a), then the origin of CSK can be traced back to cyanobacteria. If the perceived signalling role of CSK (chapter 5) is overridingly important in the maintenance of chloroplast genomes in evolution, then CSK is predicted to have a wide distribution in photosynthetic eukaryotes. These two predictions are tested with the available genome sequences of model eukaryotic and prokaryotic organisms.

7.2 Orthologues of CSK are present in all major lineages of photosynthetic eukaryotes

Arabidopsis CSK was used to search against genomic and EST sequence databases. Sequence similarity search was performed with blastp, blastn, tblastn and psi-blast (see Chapter 2 for details of these methods). The non-redundant (nr) protein database and the EST database at National Center for Biotechnology Information (NCBI), algal and plant genomes at Joint Genome Institute (JGI), the *Cyanidioschyzon merolae* genome database, Cyanobacteria genome database at Cyanobase and the Chloroplast Genome Database at Penn State University (Chloroplast DB) were used to search for CSK homologues.

It has been noted that the kinase domains of any two sensor histidine kinases typically share 20 – 50% sequence identity with an average identity of 25% (Chang

and Stewart, 1998). This means that there can arise sequence homology between any two histidine kinases simply because they share the conserved kinase domain characteristic of histidine kinase family of proteins. This sequence homology in the kinase domain however does not qualify them as 'true' orthologues. The true homology between two histidine kinases should extend to their sensor domains, which are unique for every class of sensor histidine kinases. When searching for putative CSK orthologues, this notion of homology was taken into account.

Full-length CSK orthologues can be identified in *Oryza sativa*, *Populus trichocarpa*, *Vitis vinifera*. Partial CSK orthologues can be found in other higher plants such as *Nicotiana*, *Citrus, Zea mays, Gossipium* and many others. Partial CSK orthologues are also found in gymnosperms such as *Pinus taeda* and *Picea glauca*. A full-length CSK orthologue can be readily identified in the recently sequenced moss, *Physcomitrella patens* (Rensing et al., 2008). Full-length CSK orthologues are also seen in the smallest free-living eukaryote, *Ostreococcus tauri* (Derelle et al., 2006) and in the ancient red alga, *Cyanidioschyzon merolae* (Matsuzaki et al., 2004). Full-length CSK genes are also seen in the genome sequence of the diatoms *Phaeodactylum tricornutum* (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) and *Thalassiosira pseudonana* (Armbrust et al., 2004).

7.3 CSKs share a common ancestor with the cyanobacterial histidine sensor kinase, Hik2

Sequence similarity searches further revealed that, apart from the plant and algal orthologues of CSK, the immediate bacterial homologues of CSK are to be found in cyanobacteria (Figure 7.1). The cyanobacterial CSK homologue, annotated as histidine kinase 2 (hik2) (Figure 7.2) in *Synechocystis* PCC6803, has orthologues in all the 30 or so complete genome sequences of cyanobacteria. Cyanobacterial homologues of CSK from species such as *Anabaena* sp. PCC 7120 show homology to the entire CSK sequence including the GAF domain (Figure 7.2). The homology,

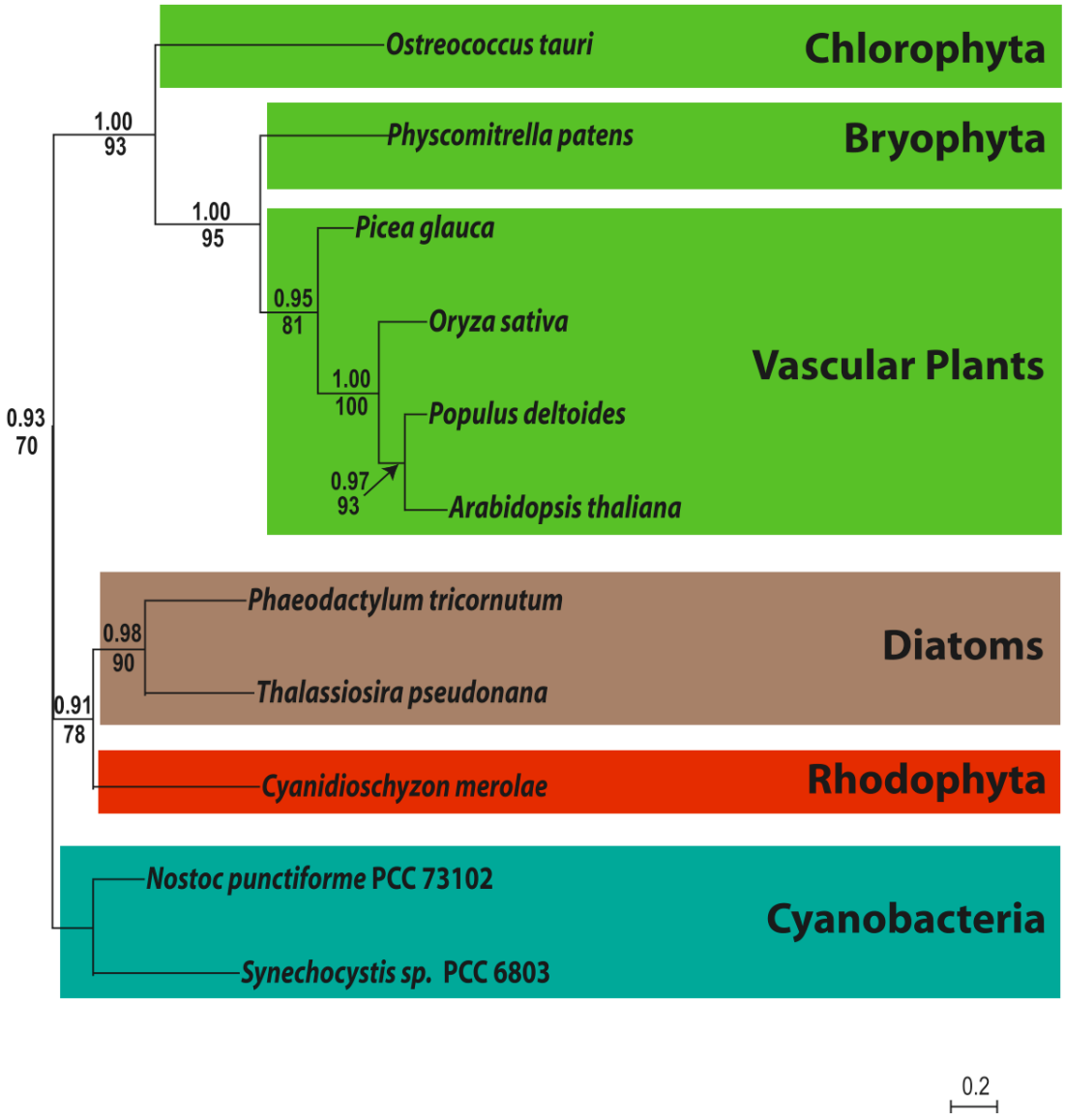


Figure 7.1 Chloroplast Sensor Kinase (CSK) is present in all major plant and algal lineages and evolved from a Hik2-like cyanobacterial histidine sensor kinase. Bayesian phylogeny of CSK is derived from the conserved kinase domain shared by all CSK orthologues. Posterior probabilities are shown above nodes and PHYML 2.4.5. bootstrap values are shown below nodes. Evolutionary-distance scale is shown at the bottom of the phylogenetic tree. A distance of 0.2 units represents 20 substitutions per 100 nucleotides. (Figure kindly provided by Matthew B. Rogers).

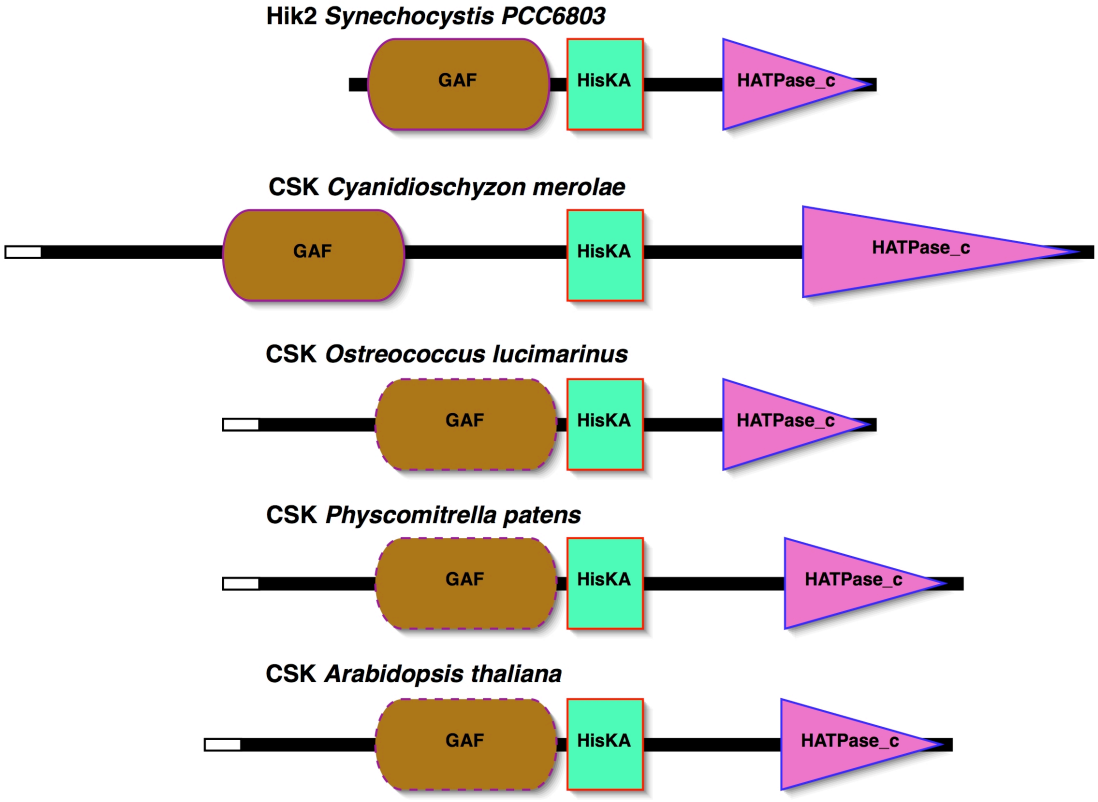


Figure 7.2 Conserved functional domains in CSK and representative algal and cyanobacterial homologues (redrawn from SMART database predictions). Domain denotations: HATPase_c, ATP-binding domain; HisKA, site of histidine auto-phosphorylation and dimerisation; GAF, sensor domain. For the *Ostreococcus* and the *Physcomitrella* CSKs, the predicted GAF domain is below the curated threshold score of SMART database. For the *Arabidopsis* CSK, SMART database does not predict a GAF domain, even though the *Arabidopsis* CSK sensor domain shows significant sequence homology to the GAF domain of cyanobacterial Hik2. Domain boundaries of GAF domains that fall below the curated threshold score, or that are not predicted by SMART database, are shown in broken lines. The chloroplast-targeting signal is represented as a white rectangle at the N-terminus.

however, does not extend to the N-terminal region of *Arabidopsis* CSK. The N-terminal region of CSK includes the chloroplast transit peptide, which was presumably acquired later in CSK evolution. The cyanobacterial homology of CSK becomes more evident, especially in the GAF domain region, if the sequence search is performed with CSK sequences from lower plants and algae (Figure 7.2). This pattern may reflect the obvious sequence divergence accepted in current evolutionary models.

7.4 Discussion

The predicted cyanobacterial ancestry and the wide distribution of CSK in photosynthetic eukaryotes are borne out by the available genomic data. Presence of CSK homologues in all major lineages of photosynthetic eukaryotes suggests deep functional and evolutionary significance. The CSK gene is retained, as *Hik2*, as one of only four histidine sensor kinases in the minimal cyanobacterial genome of *Prochlorococcus marinus* SS12 (Dufresne et al., 2003). CSK is also retained in the minimal eukaryotic nuclear genomes of *Ostreococcus tauri* (Derelle et al., 2006) and *Cyanidioschyzon merolae* (Matsuzaki et al., 2004) (in which CSK is the only sensor kinase). Furthermore, CSK has persisted through a secondary symbiotic event in the ancestor of diatoms such as *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* where the chloroplast is derived from a eukaryotic, red algal symbiont, itself a product of a primary endosymbiosis involving a cyanobacterium (Gibbs, 1981). It should be emphasised that all the CSK homologues identified so far in different photosynthetic eukaryotes are products of nuclear genes and that a plastid-encoded CSK is yet to be demonstrated. This observation points to the possibility that the *CSK* gene have been relocated to the nuclear genome very early in evolution.

Even though the distribution of CSK is very wide in phylogenetic scope, it seems to miss out the description “universal” as orthologues of CSK cannot be identified in some non-green algal taxa. The apparent absence of CSK is further tested in the recently sequenced complete nuclear genome of the haptophyte,

Emiliania huxleyi (http://genome.jgi-psf.org/Emihu1/Emihu1.home.html) and in the EST database of selected non-green algal species (NCBI). In the non-green algae tested, no CSK orthologue is found either in their nuclear genome or in their plastid genome. Nevertheless, these non-green algal species, where CSK is absent, seem to encode the ycf26 sensor in their plastid genome (Duplessis et al., 2007).

As discussed in chapter 3, the phylogenetic distribution of plastid-encoded ycf26 is limited to non-green algal species and a ycf26 homologue is completely unknown in chloroplasts of the green lineage either as a chloroplast or as a nuclear gene product. On the other hand, CSK is present in green algae and plants and in those non-green algal species, which do not have the ycf26 sensor. This distribution of ycf26 and CSK seems to suggest that each of these chloroplast sensors complements the loss of the other in different lineages of photosynthetic eukaryotes. This presumed phylogenetic complementation would require that these two chloroplast sensors, CSK and ycf26, functionally replace each other. Although this possibility remains to be tested, some commonalities in the design of their sensor domains suggest a functional redundancy. The sensor domain of both CSK and ycf26 shows redox sensing sequence features (chapter 5 and chapter 9). This presumed functional redundancy between these two chloroplast sensors is further corroborated by studies involving the cyanobacterial homologues of CSK and ycf26. It emerged from these studies that there exists a functional overlap in the cyanobacterial homologues of CSK and ycf26.

The cyanobacterial homologue of ycf26, Hik33, is known to regulate photosynthetic gene expression in response to nutrient stress and high light intensities (Hsiao et al., 2004). Hik33 is also known in cyanobacteria as a cold sensor (Mikami et al., 2002), as a sensor of osmotic stress (Mikami et al., 2002, Shoumskaya et al., 2005, Paithoonrangsarid et al., 2004), and as a sensor involved in the expression of oxidative stress-inducing genes . Hik2, the CSK homologue in cyanobacteria, has also been implicated in sensing osmotic stress (Paithoonrangsarid et al., 2004) and low temperature (Mikami et al., 2004). It has been further proposed

that hik2 and hik33 are involved in the tolerance of PS II to environmental stress (Mikami K, 2003).

Another interesting absence of a CSK homologue, apart from the ones discussed for non-green algae, could be in the recently sequenced *Chlamydomonas reinhardii* genome (Merchant et al., 2007). Sequence similarity searches cannot find any CSK homologues in the *Chlamydomonas* genome, yet it may be premature to conclude that CSK is absent in *Chlamydomonas*. A genome may remain incomplete due to flaws associated with the automated annotation of sequences; some genes may be missed out in otherwise "completed" genome. Additionally, the presence of a CSK homologue in *Ostreococcus* demonstrates the fact that CSK is not entirely absent in the chlorophycean algae.

If it turns out that CSK is indeed absent in *Chlamydomonas*, could there be another sensor kinase in its chloroplasts, and if there is one, can it compensate for the absence of CSK? These are open questions! An interesting possibility is to check whether the retinal protein, chlamyopsin (Deininger et al., 1995), plays any role in the regulation of *Chlamydomonas* chloroplast gene expression. Chlamyopsin proteins have histidine kinase domains and we find that the chlamyopsin 5 protein shows some homology with the ycf26 sensor kinase of non-green algae (data not shown). The widely thought functional role of chlamyopsin as a photoreceptor for phototaxis and photophobic responses has been recently challenged in *Chlamydomonas* (Fuhrmann et al., 2001). Chlamyopsin isoforms can be detected in some chloroplast fractionation studies (Yuichiro Takahashi, personal communication) although contamination from the eyespot cannot be ruled out at this stage. Further localization studies coupled with functional characterisation of chlamyopsin null mutants are required to probe the role of this interesting class of proteins in *Chlamydomonas*.

8.1 Introduction

A two-component system consists of a sensor kinase and a response regulator, acting together. The sensor kinase component detects changes in environmental conditions but it is the response regulator that ultimately brings about the appropriate physiological response. The output or effector domains of response regulators are predominantly DNA binding motifs; hence the resulting physiological responses involve transcriptional regulation. Having established a highly conserved, widely distributed chloroplast sensor kinase, and its role in transcriptional regulation of chloroplast gene expression, it is time to ask what actually mediates CSK's action in chloroplasts? The distribution and function of chloroplast response regulators is therefore extremely important not only to fully understand the action and operation of sensor kinases in chloroplasts but also to elucidate the nature and relevance of chloroplast transcriptional responses that they regulate.

Another important aspect of the chloroplast two-component system examined in this chapter is how the molecular evolution of CSK affects the evolution of its response regulator partner. The co-evolution of sensor kinases and their response regulator partners is well documented in the scientific literature (Koretke et al., 2000, Pao and Saier, 1997, Skerker et al., 2008). Molecular evolution has resulted in the replacement of the conserved histidine autophosphorylation site with a glutamate residue in plant CSKs (chapter 3). In *Arabidopsis* CSK, this change in the amino acid chemistry at the autophosphorylation site has replaced its ancestral histidine kinase activity with tyrosine kinase activity (chapter 6). A change in the phosphoryl group chemistry of the sensor kinase will have consequences for its cognate response regulator protein, because two-component systems employ a phosphoryl group transfer pathway as their signalling mechanism. Since phosphohistidines have a large negative free energy of hydrolysis, phosphoryl group transfer from the phosphohistidine of the sensor to the aspartate of the response regulator is thermodynamically allowed, but some other amino acid combination may forbid phosphoryl group transfer from the sensor to the response regulator.

This chapter will also examine what lessons can be learned from the cyanobacterial homologues of chloroplast response regulators. Some novel insights into the functional roles and relationships of chloroplast two-component systems will be gained from their cyanobacterial homologues, for which more data are available. Sequence similarity searches and databases outlined in the previous chapter will be used for analysing the distribution and function of chloroplast response regulators.

8.2 Chloroplast response regulators in non-green algae

The first report of a chloroplast response regulator came as a simple consequence of systematically sequencing chloroplast genomes. The first chloroplast response regulator was described in red algae and had a high sequence similarity to the bacterial OmpR response regulator transcription factor (Kessler et al., 1992). This chloroplast response regulator is known as ycf27, as "OmpR-like protein", or simply as "OmpR". It has been also called as transcriptional regulatory gene 1 (trg1) in *Heterosigma akashiwo* (Jacobs et al., 1999). Ycf27 is found (Table 8.1) in one or two gene copies in many non-green algal lineages such as in rhodophytes, *Porphyra purpurea*, *P. yezoensis*, *Porphyridium aerugineum*, *Cyanidioschyzon merolae*, *Cyanidium caldarium*, *Galdieria sulphuraria, Gracilaria tenuistipitata* and *Rhodella violacea*; in haptophytes, *Emiliania huxleyi*; in raphidophytes, *Heterosigma akashiwo*; in glaucophytes, *Cyanophora paradoxa* and in cryptophytes, *Guillardia theta* (Duplessis et al., 2007).

Our similarity searches additionally identify ycf27 homologues in other cryptophytes such as *Rhodomonas salina* and, for the first time, in the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* (Table 8.1). For the above diatoms, it must be emphasised that ycf27 homologues are encoded in the nuclear genome. This nuclear location of ycf27 gene in diatoms is in stark contrast with the hitherto known plastid location of this gene. The *ycf27* gene, although

Table 8.1 Distribution of chloroplast response regulators and response regulator-like proteins in photosynthetic eukaryotes

Taxonomic group/Organism	Ycf27	Ycf29	TCP34
Glaucophytes			
Cyanophora paradoxa	√*	√	-
Rhodophytes			
Porphyra purpurea	√*	√	-
Porphyra yezoensis	√*	√	-
Porphyridium aerugineum	√*	-	-
Cyanidioschyzon merolae	√*	√	×
Cyanidium caldarium	√*	√	-
Galdieria sulphuraria	√	-	-
Gracilaria tenuistipitata	√*	√	-
Rhodella violacea	√*	-	-
Haptophytes			
Emiliania huxleyi	√*	√N	×
Raphidophytes			
Heterosigma akashiwo	√*	-	-
Cryptophytes			
Guillardia theta	√*	√	-
Rhodomonas salina	√	√	-
Bacillariophytes			
Phaeodactylum tricornutum	√N	√N	×
Thalassiosira pseudonana	√N	√N	×
Viridiplantae			
Chlorokybus atmophyticus	√*	-	-
Ostreococcus tauri	×	×	√N
Ostreococcus lucimarinus	×	×	√N
Physcomitrella patens	×	×	√N
Arabidopsis thaliana	×	×	√*N

The tick (√) indicates the presence and the cross (×) indicates the absence of ycf27/ycf29/TCP34. The dash (-) indicates that the complete nuclear genome sequence for that taxon is not available, so the presence or absence of ycf27/ycf29/TCP34 is unknown. The superscript (*) denotes that the occurrence is based on earlier reports (Duplessis et al., 2007, Weber et al., 2006). Chloroplast response regulators are mostly chloroplast gene products, but if chloroplast response regulators exist as nuclear gene products, the nuclear location of the corresponding gene is indicated by the

superscript (^{N}). The taxonomic group "Viridiplantae" means "Green Plants", and includes green algae, lower and higher plants.

widely distributed in non-green algae, is notably absent in the chloroplast genome of the green lineage. Amongst more than 75 sequenced chloroplast genomes of green algae and land plants, only the charophyte *Chlorokybus atmophyticus* has the *ycf27* gene (Duplessis et al., 2007).

The predicted domain architecture of ycf27 conforms to known domain features of response regulators. The receiver domain is typical in having all three sequence motifs (Figure 8.1), including the invariable aspartate residue that receives phosphate from the histidine residue of sensor kinases. The effector domain of ycf27 is a winged helix-turn-helix motif. This winged helix-turn-helix is a variation of the well-known DNA binding helix-turn-helix domain. In addition to binding DNA, the winged helix-turn-helix motifs also interact with RNA polymerases to influence transcriptional regulation of genes (Martinez-Hackert and Stock, 1997).

Ycf29 is another chloroplast response regulator in the non-green algae, also identified as a by-product of chloroplast genome sequencing. This chloroplast response regulator shows high sequence similarity to the NarL response regulator transcription factor of *E. coli*. Ycf29 is also known as tctD in *Gracilaria tenuistipitata* or as NarL-like transcriptional regulator. Ycf29 often co-occurs with ycf27 in some but not all non-green algae (Table 8.1). Like ycf27, the phylogenetic distribution of ycf29 is also limited to non-green algae. Our analysis (Table 8.1) finds Ycf29 homologues to be present in rhodophytes, *Porphyra purpurea*, *P. yezoensis*, *Cyanidioschyzon merolae*, *Cyanidium caldarium*, *Gracilaria tenuistipitata*; in glaucophytes, *Cyanophora paradoxa* and in cryptophytes, *Guillardia theta*, *Rhodomonas salina*. We also identified nuclear-encoded ycf29 homologues in diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*

and in the haptophyte, *Emiliania huxleyi*. Again this nuclear location contrasts with the known plastid location of this gene in other algal species.

Ycf29 protein is very similar to ycf27 response regulator, apart from a small difference in their respective effector domains. The receiver domain of ycf29 has all the characteristic motifs of response regulators (Figure 8.1). Ycf29 does not have the extended wing structure in its DNA–binding helix-turn-helix domain.

8.3 Chloroplast response regulators in green algae and plants

Chloroplasts in the green lineage, with the single exception of a ycf27 homologue in *Chlorokybus atmophyticus*, do not seem to encode response regulators (Duplessis et al., 2007). This observation does not preclude the existence of chloroplast response regulators in green algae and plants as products of nuclear genes, as shown here for chloroplast sensor kinases (chapter 3). Homologues of ycf27 or ycf29 cannot readily be identified in sequenced nuclear genomes of green algae and plants. Nevertheless, it is premature to conclude that nuclear-encoded ycf27 and ycf29 homologues are absent from chloroplasts of green algae and land plants as yet. One possibility is that the chloroplast response regulators in green lineages are modified so as to accommodate input from a modified histidine kinase such as CSK. One or more modified response regulators might then evade identification by conventional sequence similarity searches.

Along these lines, a nuclear-encoded, modified response regulator protein, TCP34, has been identified in higher plant chloroplasts (Weber et al., 2006). This protein seems to be conserved in all sequenced plant genomes. Our analysis (Table 8.1) identifies a homologue of TCP34 in the moss, *Physcomitrella patens*. *Physcomitrella* also has a paralogue of TCP34 (Andrew Cuming, University of Leeds, personal communication). We also identify (Table 8.1) single homologues of TCP34 in the nuclear genome of chlorophycean algae, *Ostreococcus tauri* and *Ostreococcus lucimarinus*, but interestingly, not in *Chlamydomonas reinhardii*. The

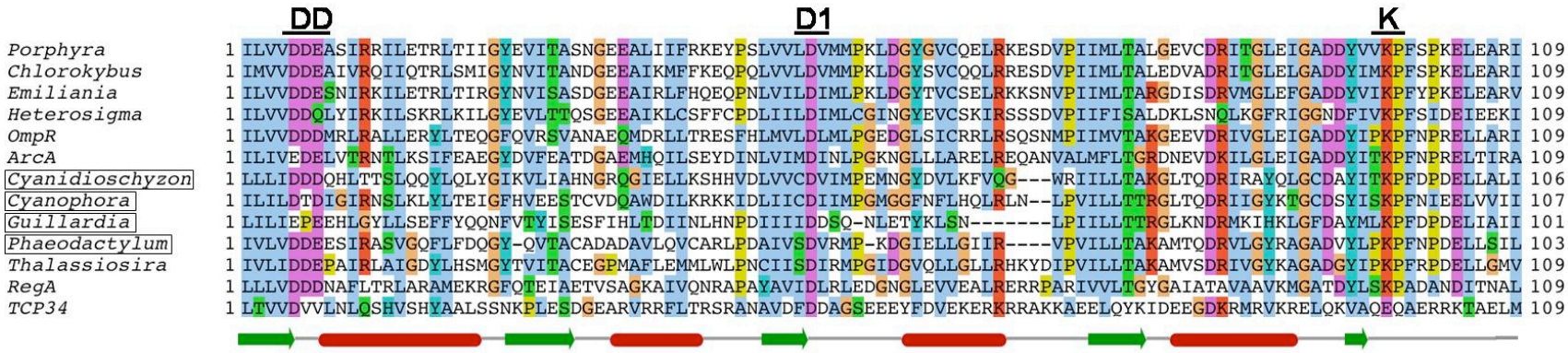


Figure 8.1 Conserved sequence features of chloroplast response regulators. Amino acid sequences corresponding to the receiver domain of representative ycf27, ycf29 and TCP34 proteins are aligned with those of three canonical response regulators in bacteria, ArcA, RegA and OmpR. Species names of ycf29 sequences are boxed. The residue colouring scheme is same as in Figure 3.1. Predicted secondary structures are shown at the bottom. Conserved sequence motifs of the receiver domain, DD, D1 and K are indicated on top of the alignment. The fully conserved D1 motif, which lies in the loop connecting the third *beta*-sheet and the third *alpha*-helix, is the conserved aspartate residue that receives the phosphate group from the phosphohistidine of the sensor kinase. Arabidopsis TCP34 shows a moderately conserved receiver domain with the 'K motif' still being present but not entirely aligned with those from other response regulators.

moss and algal TCP34 homologues are also localised in the chloroplast, according to sub-cellular prediction programs (results not shown).

The TCP34 protein combines a moderately conserved receiver domain with a tetratricopeptide repeat (TPR) as an effector domain (Weber et al., 2006). It also seems that a part of the receiver domain of this protein may be co-opted as a DNA–binding, helix-turn-helix motif.

8.4 Discussion

Although the DNA binding winged helix-turn-helix motif in the output domain of the ycf27 response regulator suggests a transcriptional regulatory role, not much is known regarding the function of this protein in non-green algae. This is partly due to difficulties in generating knockout mutant lines in these algal groups (Duplessis et al., 2007). A transcriptional regulatory role for ycf27 is assumed in the red alga *Cyanidioschyzon merolae*, where it is shown to bind upstream regions of *psbD/C* operon and the *psbA* gene (Minoda and Tanaka, 2005). The binding of ycf27 to *psb* gene promoters is thought to be part of an acclimatory response to high light intensity.

It has been recognized that the ycf27 has a cyanobacterial homologue known as response regulator 26 (rre26). Rre26 has been also called rpaB (for regulator of phycobilisome association) (Ashby and Mullineaux, 1999). A closely related protein, rpaA, probably a paralogue of rpaB, is also present in cyanobacteria (Ashby and Mullineaux, 1999). These two response regulators, rpaA and rpaB, are involved in synthesis of factors responsible for coupling phycobilisomes to photosystems (Ashby and Mullineaux, 1999). RpaA response regulator has been further shown to be part of the output pathway of the central circadian oscillator in cyanobacteria (Takai et al., 2006). It should be made clear that all the ycf27 response regulators of non-green algae seem to be derived from the rpaB response regulator of cyanobacteria and not from the rpaA paralogue. RpaB has been recently shown to bind the promoter region

of the high-light-inducible *hliB* gene in *Synechocystis* PCC 6803 (Kappell and van Waasbergen, 2007).

In *Synechocystis* sp. PCC 6803, various functional and mutagenesis studies, together with high-throughput two-hybrid screening, have identified cognate pairs or interaction partners in different two-component systems (Sato et al., 2007). These studies find rre26 to be the response regulator of hik33, the cyanobacterial homologue of ycf26. The action of this multi-sensor kinase (discussed in previous chapter), hik33, is presumably mediated through the rre26 response regulator (Ashby and Houmard, 2006).

The DNA binding and transcriptional regulatory properties of ycf29 is demonstrated in *Cyanidioschyzon merolae*, where it is shown to bind genes encoding the components of the phycobilisome proteins such as cpcA and apcE (Minoda and Tanaka, 2005). Transcriptional activation of these genes by ycf29 is believed to be part of an acclimatory response to decreased light intensity. A cyanobacterial homologue of ycf29 is known to exist in the form of response regulator 1 (rre1). In cyanobacteria, rre1 is the cognate response regulator of hik2, the CSK homologue (Sato et al., 2007). It is believed that the physiological responses governed by hik2 in cyanobacteria, as discussed in the previous chapter, are mediated through the action of rre1 response regulator (Ashby and Houmard, 2006).

TCP34, the response regulator-like protein of plant chloroplasts, has been shown to bind *psaA*, *rbcL*, *psbC* and *psbD* chloroplast gene core promoters in spinach (Weber et al., 2006). It has been also demonstrated that TCP34 is phosphorylated. TCP34 does not appear to have cyanobacterial counterparts. The phylogenetic signature of this protein seems to have been lost during its evolution. TCP34 might represent a eukaryotic innovation in which the receiver domain of a response regulator is fused to a tetratricopeptide repeat, the latter motif being usually found in eukaryotes (Weber et al., 2006).

It has been suggested that the distribution of ycf27 genes is somehow correlated with the retention or loss of phycobilisomes in various algal groups (Ashby et al., 2002). The loss of ycf27 genes from plastids has been linked with the loss of phycobilisomes. But the demonstration of ycf27 genes in the non-phycobilisome containing genera (Table 8.1) *Heterosigma akashiwo*, *Guillardia theta*, *Emiliania huxleyi* and *Chlorokybus atmophyticus* proves that no such link with phycobilisomes is required for retention of ycf27 genes (Duplessis et al., 2007).

Similarly, our analysis shows that the loss of ycf29 response regulator gene is not linked with the loss of phycobilisomes as ycf29 genes are present in non-phycobilisome containing species such as the diatom *Thalassiosira pseudonana* and the cryptophyte *Guillardia theta* (Table 8.1). Nonetheless, there may exist some functional link between ycf29 and phycobilisomes. For instance, the ycf29 response regulator is involved in the transcriptional activation of phycobiliprotein genes as part of an acclimatory response to low light intensity in *Cyanidioschyzon merolae* (Minoda and Tanaka, 2005). The demonstration that ycf27 and ycf29 genes are retained in non-phycobilisome containing algae is a clear indicator that these two response regulators have other target genes in chloroplasts.

Analysis of the distribution, sequence and functional features of chloroplast response regulators implies a key role for these proteins in regulating chloroplast gene expression by means of transcriptional control. Studies involving their cyanobacterial homologues provide further evidence for regulation of photosynthetic genes by chloroplast response regulators. Nevertheless, analysis of knockout mutants is required to elucidate different aspects of the transcriptional control mechanism provided by the chloroplast response regulators. The nature and number of target genes and the physiological context in which they are regulated by chloroplast response regulators also remain to be determined.

The presence of cyanobacterial homologues certainly strengthens the genealogy of chloroplast two-component systems, but it has also uncovered the

functional relationships between chloroplast two-component systems. Two cognate pairs of two-component systems identified in *Synechocystis* sp. PCC 6803 are important from a plastid perspective. These are the cognate pairs formed between hik33 and rre26, and hik2 with rre1. This means that ycf26 and ycf27 form one cognate pair while CSK and ycf29 form another. The deviations from these ancestral combinations and its consequence for the distribution of chloroplast two-component systems are discussed in the next chapter.

Establishing that the CSK homologue, hik2 and the ycf29 homologue, rre1, form a cognate pair of sensor and response regulator in cyanobacteria (Sato et al., 2007) has unravelled an interesting but wholly unexpected aspect of CSK evolution. This functional relationship of hik2 and rre1 in cyanobacteria implies that CSK, the nuclear-encoded, predominant chloroplast sensor of the green lineage, is the cognate sensor of a plastid-encoded response regulator that is found exclusively in non-green algae. CSK is thus related to two-component systems of non-green algae by descent and by a shared functional role. The functional roles of CSK established here (Chapter 5) can then be securely transferred to chloroplast two-component systems of the non-green algae. The same holds for assigning the known functional properties of non-green algal chloroplast two-component systems to CSK.

The 1.5 billion years of evolution that followed the symbiotic origin of chloroplast has clearly tinkered with CSK and its cognate response regulator(s). Firstly, CSK has been separated from its cognate response regulator, ycf29, in terms of gene location, which for CSK is exclusively nuclear. Secondly, the functional pairing between CSK and ycf29 or ycf27, seen in red algae and diatoms, is nonexistent in the green lineage, as the “greens” do not appear to have either ycf29 or ycf27 response regulators. The same evolutionary force that moved most chloroplast genes to the nucleus may drive the nuclear relocation of CSK gene from the chloroplast (Allen and Raven, 1996), or there could be other forces at work, which we do not understand yet. The event that caused the disappearance of the

cognate response regulator(s) could have been driven by the molecular evolution in the kinase domain of CSK.

Cognate response regulator(s) of CSK are only seen in red algae and diatoms, where CSK is an unmodified histidine kinase. When CSK becomes a modified histidine kinase, as in green algae and plants, no recognizable response regulator partners can be identified. It is known that when histidine kinases lose their histidine kinase activity, they may also lose their cognate response regulators (Koretke et al., 2000). Whether the cognate response regulator(s) of CSK are really 'lost' in the green lineage or simply become 'modified' remains to be seen.

TCP34, the modified response regulator protein of "green" chloroplasts thus fits nicely into the scheme of regulation necessitated by the molecular of evolution of CSK. Our sequence analysis suggests a greater variation and consequently less sequence homology (Figure 8.1) in the receiver domain of TCP34 from the green algae and plants than in the receiver domain of response regulators of non-green algae. It is not yet clear whether this amounts to a co-evolution in TCP34 to accommodate a CSK-like sensor, which has similarly evolved in its kinase domain in green algae and plants. But this is what one would expect if TCP34 were indeed the response regulator partner of CSK in green chloroplasts. Before proceeding with this or any other possibility, the response regulator credentials of TCP34 in chloroplasts should be settled.

9.1 Summary of the main findings presented in this thesis

- A nuclear-encoded two-component sensor kinase is present in algal and plant chloroplasts.
- This Chloroplast Sensor Kinase (CSK) is now identified and characterised.
- CSK shows molecular evolution with respect to the histidine residue in the autophosphorylation site. Non-green algal CSKs exist as unmodified sensor histidine kinases, while green algal and plant CSKs are modified sensor histidine kinases.
- CSK is localized in the chloroplast and is likely to be a stromal protein.
- CSK is imported into chloroplasts from the cytosol as a precursor that seems not to be processed proteolytically after import.
- CSK functions as the sensor component in a redox-responsive two-component system that connects photosynthetic electron transport to gene transcription in chloroplasts.
- The *Arabidopsis* CSK autophosphorylates *in vitro* and the autophosphorylation requires divalent cations Mn^{2+} or Ca^{2+}. The acid and alkali stability of the autophosphorylated amino acid(s) in CSK suggests autophosphorylation on tyrosine residue.
- CSK has a wide phylogenetic distribution, in contrast to the chloroplast-encoded ycf26 sensor kinase of non-green algae.
- Hik2 in cyanobacteria is a homologue of CSK. This finding sheds light on the cognate response regulator partner of CSK and further reveals an unexpected phylogenetic connection between CSK and chloroplast two-component systems of non-green algae.
- In the light of available data, the general functional property of chloroplast two-component systems is one of providing a regulatory coupling between photosynthesis and gene transcription in chloroplasts.
- Chloroplast two-component systems are likely to be universal in photosynthetic eukaryotes, and they persist in chloroplasts as products of nuclear genes even where chloroplast genomes no longer encode them.

- There exists an evolutionary continuity in the distribution and function of two-component systems in chloroplasts and in their cyanobacterial ancestors.
- The distribution of chloroplast two-component systems follows a lineage specific pattern.
- The endogenous rhythm of mRNA accumulation in chloroplasts is generated by the action of chloroplast two-component systems and that a chloroplast location for chloroplast two-component system genes has been favoured in evolution to fulfil this functional role.

9.2 Chloroplast two-component systems: the missing link between photosynthesis and gene expression

Analysis of the functional properties of chloroplast two-component systems reported in this thesis suggests an indispensable role for these proteins in linking photosynthetic electron transport to the transcription of genes encoding the photosynthetic machinery. Chloroplast two-component systems fulfil this vital role by acting as redox sensors and redox response regulators (Allen, 1993a, Allen, 1993b) inside the chloroplast. Chloroplast sensor kinases, acting as redox sensors, monitor the flux of electron transport in the thylakoid membrane. The flux of electron transport lies at the heart of photosynthetic activity. Sensor kinases pass this information on to response regulators, which, when required, selectively turn photosystem genes on and off. This feedback regulation brings about redox homeostasis and photosynthetic activity in chloroplasts that becomes rapidly and unconditionally optimised to changing environmental conditions.

Although the functional role of CSK has been worked out to a significant extent, little is known concerning the role of the other chloroplast sensor, ycf26, which is predominantly found in non-green algae. This is partly because knock-out mutant lines are not available in these algae (Duplessis et al., 2007). Nonetheless, a transcriptional regulatory role of photosynthesis-related genes is what the cyanobacterial homologue of ycf26 seems to indicate (chapter 7). Additionally the

sensor domain of ycf26, like that of CSK (chapter 5), is suggestive of a redox-sensing function.

The ycf26 sensor kinase seems to be a transmembrane protein as most of its examples have 2-3 predicted transmembrane helices. The predicted topology (Figure 9.1) of ycf26 includes one or two transmembrane helices at the N-terminus and a lumenal loop of 120-130 amino acids followed by another transmembrane helix. The predicted stromal part of the enzyme has an HAMP linker domain and a PAS domain followed by the kinase domain (Figure 9.1). HAMP domains are known to act as linkers that connect the periplasmic sensor domain with the cytoplasmic kinase domain of transmembrane histidine kinases (Aravind and Ponting, 1999). PAS domains are well known for their role as sensors of redox state (Taylor and Zhulin, 1999). It appears that in most ycf26 proteins, the predicted thylakoid lumenal loop and the internal PAS domain act as two separate sensor domains (Morrison et al., 2005). Ycf26, with its lumenal loop and PAS domain acting as two separate sensor domains, will be able to integrate multiple signals and act as a multi-sensor (Morrison et al., 2005). However, the redox-sensing functional role of ycf26 is underscored by the observation that this sensor is seen in its minimal form, with only the PAS sensor domain and the kinase domain being present, in the raphidophycean alga, *Heterosigma akashiwo* and the cryptophycean alga, *Rhodomonas salina*.

Direct evidence for the transcriptional regulatory role of chloroplast response regulators is also lacking because of difficulties in generating knockout mutants in non-green algae (Duplessis et al., 2007). Nonetheless, a role for these proteins in photosynthetic gene transcription can be inferred from observations of plastid gene promoter binding by chloroplast response regulators and from studies involving the cyanobacterial homologues of chloroplast response regulators (chapter 8). With this being the case for chloroplast response regulators in the non-green lineages, the functional characterisation of any chloroplast response regulators in the green lineage will have to wait for their experimental identification.

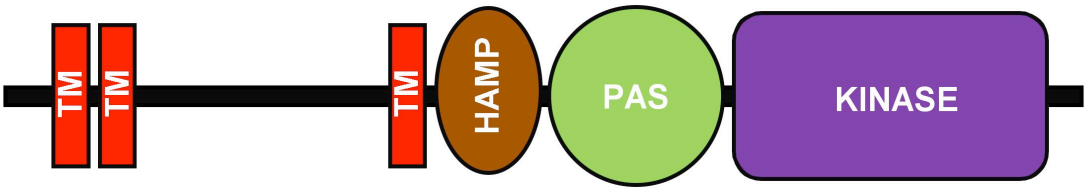


Figure 9.1 Predicted domain composition of ycf26. Individual domains are labelled as TM, Transmembrane; HAMP, domain present in histidine kinases, adenylyl cyclases, methyl-accepting proteins and phosphatases; PAS, domain named after three proteins it occurs in, per, arnt, and sim; KINASE, kinase domain of sensor histidine kinases.

Two-component systems connect photosynthesis to gene expression, but how exactly does this connection bring about homeostasis in chloroplast function? The answer lies in the chloroplast's resulting ability to control biogenesis of the electron transfer complexes participating in the light reactions of photosynthesis (Allen, 2003b). These electron transfer complexes are multi-subunit pigment-protein assemblages with their core protein components being encoded by the chloroplast genome and the peripheral protein components encoded by the nuclear genome (Figure 9.2). The biogenesis of these electron transfer complexes involves a hierarchical assembly in which the chloroplast-encoded core components are made and inserted into the thylakoid membrane first, followed by the assembly of the nuclear-encoded peripheral components (Choquet and Vallon, 2000).

Some general principles have been recognized in the highly concerted assembly processes of electron transfer complexes in chloroplasts (Choquet and Vallon, 2000). Nuclear-encoded subunits are found to be less essential to the stability of the complexes than the chloroplast-encoded core subunits. Among chloroplast-encoded subunits, there exists a sequential and ordered assembly, which arises from a hierarchical organization in the expression of these subunits. This in regulatory terms means that the translation of certain subunits is controlled by the state of assembly of the complex. And to be precise, the rate of translation of some subunits depends on other subunits being present in the membrane. The former subunit is called a CES (Control by epistasy of synthesis) protein and the latter is a dominant assembly factor (Choquet and Vallon, 2000). The dominant assembly factors are less stable in the complexes and, in the absence of their assembly partners, are rapidly degraded by proteases. CES proteins on the contrary are more stable, but their synthesis is reduced in the absence of dominant assembly factors, by an autoregulatory translational loop. In *Chlamydomonas*, mutant studies have revealed that in each complex of the thylakoid membrane, at least one subunit is under epistatic control of synthesis and thus be a CES protein. Cytochrome *f* in Cytochrome b_6f, psaA protein in PS I, D1 and CP47 in PS II, α subunit in ATP

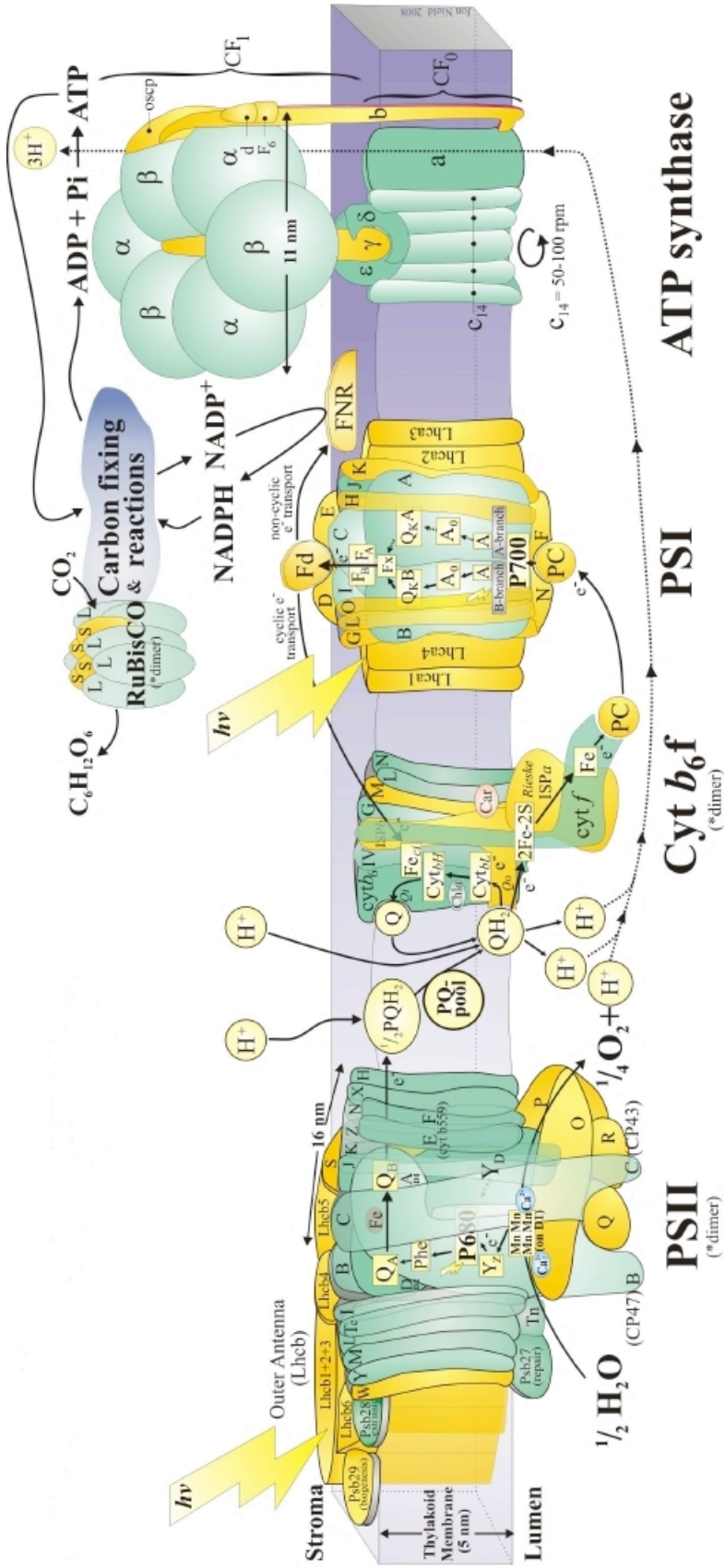


Figure 9.2 Thylakoid membrane is a genetic chimera. The dual genetic origin of all four major electron transfer complexes, PS II, Cyt b6f, PS I and ATP synthase, are shown. Chloroplast-encoded, core subunits of the electron transfer complexes are coloured in green or shades of green, while yellow or shades of yellow represents nuclear-encoded, peripheral subunits. Arrows indicate direction of electron and proton transfer through the complexes. This figure is based on the thylakoid protein composition of *Arabidopsis thaliana*. Mapping of some minor subunits and representation of their genetic location are not complete in this image. (Unpublished, figure kindly provided by Dr. Jon Nield, http://www.queenmaryphotosynthesis.org/nield/downloads.html).

synthase and the large subunit of RuBisCO are CES proteins (Choquet and Vallon, 2000).

The Control by epistasy of synthesis and the biogenesis of electron transfer complexes in general offer some important insights into the gene content of chloroplast genomes and the genetic control mechanism of photosynthetic genes. The gene content of chloroplast genomes can be explained only when the CORR hypothesis (Box 9.1) is viewed in the context of the biogenesis and assembly processes of electron transfer complexes. A synthesis of CORR (Box 9.1) and biogenesis principles (Box 9.2) might explain why species-specific differences in the assembly process could result in species-specific differences in the gene content of chloroplasts. For example, some photosystem genes such as *psaD psaE* and *psaF* are always chloroplast-encoded in red algae, but they are nuclear-encoded in green algae and plants.

Autoregulation of translation of CES proteins becomes possible only when their genes are present in the same location where the assembly takes place (Choquet and Vallon, 2000). CES properties certainly explain why genes encoding CES proteins should be there in chloroplasts, but they do not explain why genes encoding dominant assembly factors should also be there in chloroplasts. The retention of these genes, however, has to do with the nature of the biogenesis process itself, a process in which the chloroplast-encoded core subunits act as “seeds” or “nuclei” on which the nuclear-encoded peripheral subunits assemble to form functional electron transfer complexes.

The pivotal role of chloroplast-encoded subunits in assembly is further highlighted by the observation that some small subunits are encoded in the chloroplast genome and integral to assembly without forming part of a mature, functional complex (Boudreau et al., 1997). The pre-eminence of chloroplast-encoded subunits in the assembly process means that, if chloroplasts can rapidly control the availability of core-subunits, they can “make” or “break” whole electron

Box 9.1

The CORR hypothesis (Co-location for Redox Regulation) states that genes have been retained in bioenergetic organelles to provide direct and unconditional regulation of their expression by the redox state of their gene products, the core proteins of energy transduction in photosynthesis and respiration. Some important axioms and predictions of CORR are:

- **The principle of selective value of redox control**

For each gene under direct redox control, it is selectively advantageous for that gene to be retained and expressed only within the organelle.

- **The principle of unselective gene transfer and protein import**

Gene transfer from the organelle to the nucleus of the host cell is not selective for particular genes and there is no barrier to the successful import into the organelles of any precursor protein.

- **The principle of selective value for nuclear location of genes**

It is selectively advantageous to relocate those organellar genes, which do not encode core-subunits of electron transfer complexes and whose expression need not be redox-regulated, to the nucleus of the host cell.

- **The principle of primary involvement in energy transduction**

Those genes for which direct redox control is vital, and whose gene products form the core subunits of electron transfer complexes and participate in primary electron transfer reactions, are always contained in the organelles.

- **The principle of secondary involvement in energy transduction**

Genes whose products constitute the organelle genetic system, or whose products are associated with secondary events in energy transduction, may be retained in organelles in one group of organisms, but not in another.

- **The principle of nuclear encoding of redox signalling components**

The redox-signalling components, upon which the co-location of genes and gene products is based, are themselves not involved in primary electron transfer, and so their genes have been relocated to the nucleus.

transfer complexes. The precisely timed delivery of assembly-competent, chloroplast-encoded, core subunits is essential not only for the assembly process itself but also for the prompt binding of chlorophylls and other cofactors that are synthesized in the chloroplast. The deleterious effects of unbound chlorophylls and their biosynthetic intermediates are well known.

Box 9.2

The following properties of the biogenesis process of electron transfer complexes may act in conjunction with CORR to retain genes in chloroplasts,

- Electron transfer complexes in chloroplasts follow a hierarchical assembly in which the chloroplast-encoded core subunits acts as "seeds" or "nuclei" around which the nuclear-encoded peripheral subunits assemble.
- Nuclear-encoded subunits are less essential to the stability of the complexes than the chloroplast-encoded core subunits.
- Translation of certain chloroplast-encoded subunits is regulated by the state of the assembly of the complex. This regulation requires their genes being present in the same compartment where the assembly takes place.
- Some subunits that are not part of a mature, functional complex but are integral to the assembly process are encoded in the chloroplast.
- Species-specific differences in assembly can result in species-specific differences in chloroplast gene content.

Control over the availability of chloroplast-encoded core subunits can be achieved by means of both transcriptional and translational regulation, although the former has been considered as the principal mode of gene regulation. Since chloroplast-encoded core subunits comprise of both dominant assembly factors and CES proteins, it is interesting to ask whether both classes of genes are under

transcriptional control. If both categories of genes are under transcriptional control, will they differ in their levels of response, for example one class of genes respond faster than the other? The polycistronic nature of many chloroplast mRNAs, however, suggests that both dominant assembly factor and CES protein genes could be equal targets of transcriptional regulation.

If rapid regulation of chloroplast-encoded genes is what tips the equilibrium of complex formation, how will chloroplasts know which chloroplast genes should be turned on and which turned off? As discussed in chapter 1, the redox state of some photosynthetic electron carriers is a useful source of information for deciding the composition of thylakoid membranes. But, this will require a continuous dialogue between electron carriers and chloroplast genes encoding core-subunits. A minimal signal transduction chain in the form of a two-component system makes this dialogue possible (Figure 9.3). The chimeric nature of the electron transfer complexes further requires that the dialogue within the chloroplast be overheard by the nucleus, so that the availability of the nuclear-encoded peripheral subunits is also, eventually, brought under control (Figure 9.3). Plastid-to-nuclear signalling pathways (Koussevitzky et al., 2007, Fey et al., 2005a, Gray et al., 2003, Lopez-Juez and Pyke, 2005, Lopez-Juez, 2007) must have originated to achieve this level of coordination required for the assembly of electron transfer complexes.

9.3 Evolutionary continuity in the distribution and function of chloroplast two-component systems

An important prediction of the CORR hypothesis is that redox-signalling devices that connect photosynthesis to gene expression operated both before, during and after the transition from cyanobacterium to chloroplast (Allen, 2003b). This prediction is borne out by the distribution and function of chloroplast two-component systems described in this thesis. As discussed in earlier chapters, cyanobacterial homologues for chloroplast two-component systems can be recognized in extant cyanobacterial

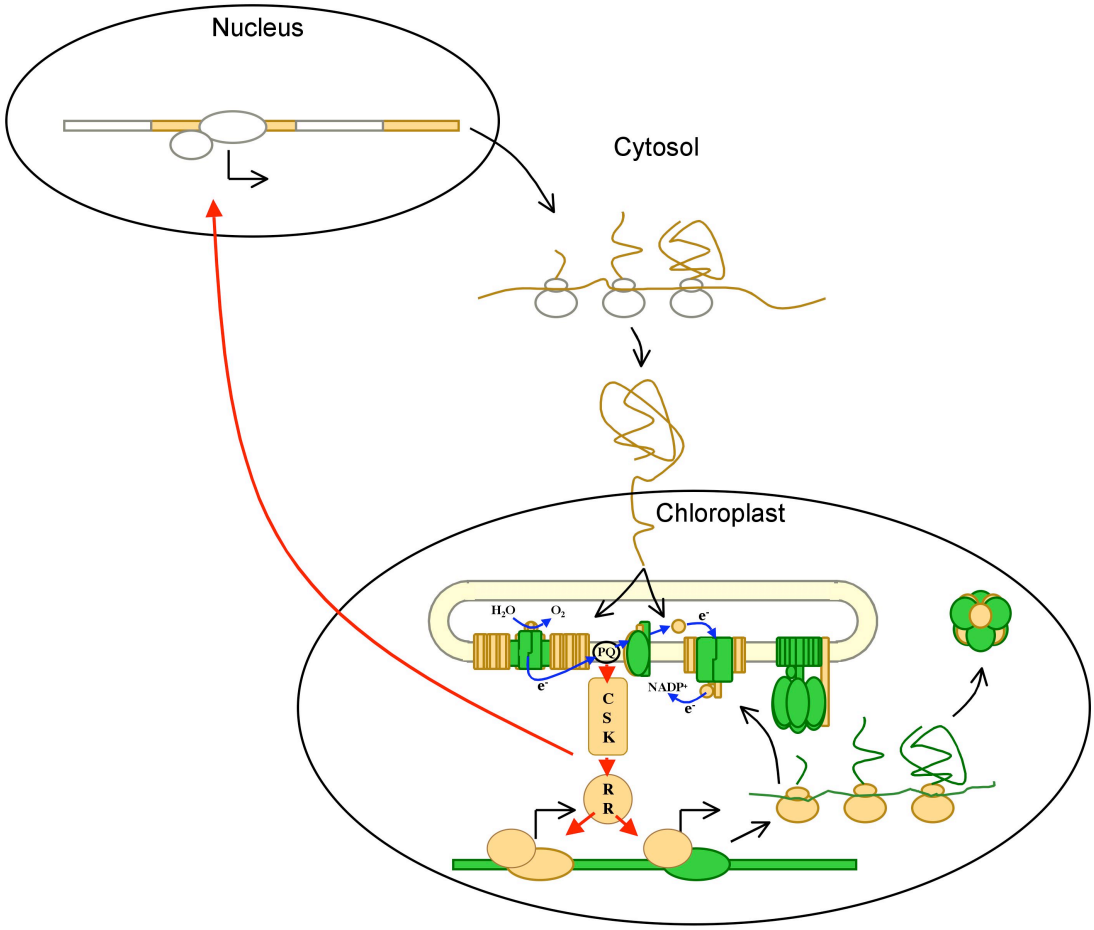


Figure 9.3 A model explaining the operation of a CSK-based two-component system in chloroplasts. CSK selectively switches on and off chloroplast genes in response to perturbations in the photosynthetic electron transport chain (depicted as electron transport from H_2O to $NADP^+$). CSK acts as a redox sensor, and reports on electron flow. A response regulator protein (RR) whose identity is yet to be revealed in green algae and plants, mediates CSK's control over gene expression. Chloroplast genes under CSK control are those that encode core components (shown in green) of the electron transfer complexes. The nuclearly-encoded peripheral components are shown in yellow. The red arrow from the chloroplast to the nucleus represents plastid-to-nuclear signalling. (Figure adapted from (Allen, 2003b)).

lineages and functional roles of the cyanobacterial counterparts are strongly suggestive of their coupling photosynthesis with gene expression. Chloroplast two-component systems thus epitomise a signalling system which could never have been put "on hold" in evolution.

An experimental strategy to test the continuity in the function of two-component systems in chloroplasts and in their cyanobacterial ancestors would be cross-kingdom complementation. If the functional continuity is correct, the cyanobacterial homologue of CSK, hik2, should be able to complement the loss of CSK; hik33, the loss of ycf26; rre26, the loss of ycf27; rre1, the loss of ycf29 in chloroplasts. For successful complementation, cyanobacterial homologues of chloroplast two-component systems will require a transit peptide for import into chloroplasts. It is, however, unlikely that hik2 would be able to complement the loss of CSK in green algae and plants, as their CSK is a modified histidine kinase.

9.4 Lineage specific retention/loss of two-component systems

The presence of cyanobacterial homologues, apart from strengthening the genealogy of chloroplast two-component systems, has uncovered the cognate interaction partners in chloroplast two-component systems. As discussed in previous chapters the two cognate pairs important from a plastid perspective in cyanobacteria are those formed between hik33 and rre26, and hik2 with rre1. These functional relationships in cyanobacteria, when extended to chloroplasts, mean that ycf26 and ycf27 form one cognate pair while CSK and ycf29 together form another. Although some chloroplasts still keep these ancestral combinations between sensors and response regulators, most chloroplasts have deviated from this pattern for reasons which we do not yet fully understand. These deviations have resulted in a striking pattern of lineage–specific retention or loss of chloroplast two-component systems (Figure 9.4).

In photosynthetic eukaryotes, chloroplast two-component systems as we understand them today consist of two sensors (CSK, ycf26), two response regulators

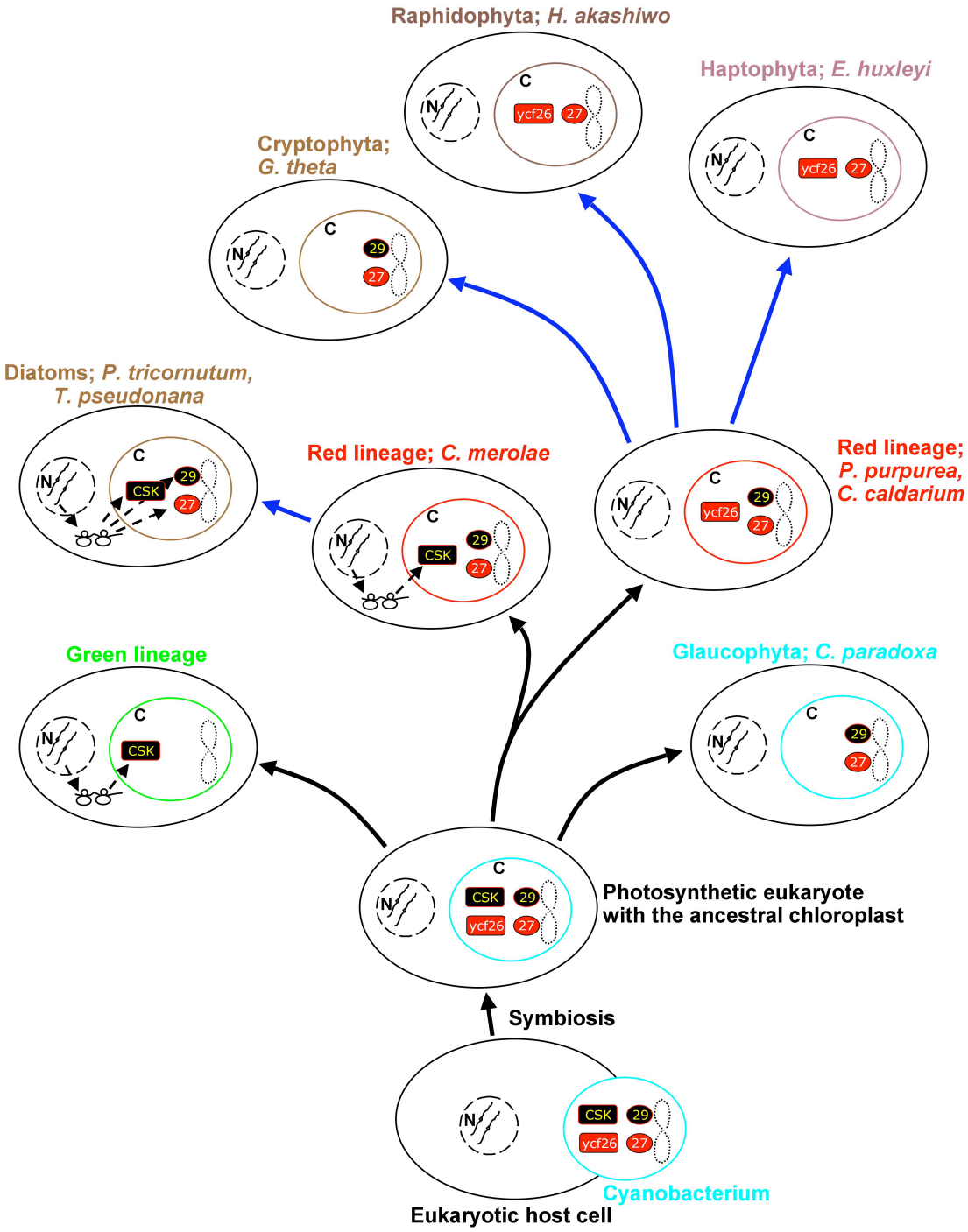


Figure 9.4 Lineage-specific distribution of chloroplast two-component systems. Cyanobacteria and chloroplasts are represented as ovals. Their names and oval representations are coloured according to their major pigments. Chloroplasts are additionally marked "C" and the nucleus is represented by dashed circles and marked "N". When genes encoding either component of the two-component

systems are moved to nucleus, synthesis of this component in the cytoplasm and import back to chloroplasts is indicated by ribosomes and dashed arrows. Lineages of non-green algae leading to diatoms, cryptophytes, raphidophytes and haptophytes have involved secondary endosymbiosis with a red algal symbiont and are represented by thick-blue arrows from the red lineage. "27" and "29" denote the non-green algal response regulators ycf27 and ycf29 respectively.

(ycf29, ycf27), and an additional response regulator like-protein (TCP34). The distribution of these proteins seems to follow a lineage–specific pattern (Figure 9.4). In the lineage of rhodophytes leading to *Porphyra purpurea*, *Cyanidium caldarium* and *Gracilaria tenuistipitata*, chloroplast two-component systems consist of the ycf26 sensor and two response regulators, ycf27 and ycf29, all chloroplast-encoded (Figure 9.4). In the raphidophyte *Heterosigma akashiwo* and the haptophyte, *Emiliania huxleyi* the same combination is seen except that the ycf29 response regulator is missing (Figure 9.4). In the ancient red algae, *Cyanidioschyzon merolae*, the response regulators are ycf27 and ycf29 proteins, both encoded in the chloroplast, while the sensor ycf26 is lost and presumably replaced by the nuclearly encoded CSK (Figure 9.4).

The glaucophyte *Cyanophora paradoxa* and the cryptophyte *Guillardia theta* resemble *Cyanidioschyzon* in their contingent of chloroplast two-component systems (Figure 9.4), but a nuclearly encoded sensor kinase is yet to be demonstrated in their chloroplasts. Chloroplasts of bacillariophytes such as *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* do not encode any two-component system, nevertheless they seem to contain chloroplast two-component systems as products of nuclear genes, since homologues of *CSK*, *ycf27* and *ycf29* genes can be identified in their genomes.

Chloroplasts in the green lineage (Figure 9.4), with the single exception of *Chlorokybus atmophyticus*, seem not to encode two-component systems. Orthologues of CSK are readily identifiable in the sequenced nuclear genomes of higher plants, in the moss *Physcomitrella patens* and in the prasinophycean alga,

Ostreococcus tauri. In contrast, the identity of chloroplast response regulators in green algae and land plants, again with the exception of *Chlorokybus atmophyticus*, remain shrouded in mystery. Whether the nuclear-encoded TCP34 protein can be counted as a genuine chloroplast response regulator remains to be seen.

Distribution of chloroplast two-component systems appears to be a phylogenetic patchwork (Figure 9.4). Nevertheless a pattern emerges with regard to the location of genes encoding chloroplast two-component systems. These genes are shown to move from chloroplast to the nuclear genomes as we proceed from non-greens to greens (Figure 9.4). Sequencing of more chloroplast and nuclear genomes may uncover yet further combinations of sensors and response regulators in chloroplasts. Genome sequencing will also answer questions such as the following. Are there nuclearly encoded ycf26 homologues in algae and plants? Will a CSK homologue be found encoded in a chloroplast?

9.5 Factors accounting for the uneven distribution of two-component system genes between chloroplasts and nucleus

The emerging properties of chloroplast two-component systems and their cyanobacterial homologues demonstrate their role in linking photosynthesis with gene expression. How does this general function account for the uneven distribution of their genes between chloroplast and nuclear genomes? One hypothesis is that the retention of two-component genes in chloroplasts confers a selective advantage to the organism bearing them, since they may then be able to mount a rapid and elaborate adaptive response to serious environmental assaults such as high light intensities (Allen, 2003b). This amplified response is thought to be mediated through a positive feedback loop in the two-component regulatory system, since its components are encoded in the operon which they, themselves, control (Allen, 1995, Allen and Nilsson, 1997). Such a positive feedback loop has been shown to promote a transcriptional surge that jump–starts virulence in *Salmonella typhimurium* (Shin et al., 2006). Amplification of signals within chloroplasts may also explain the

occurrence of chloroplast-encoded response regulator genes in more than one copy number in some non-green algae. Allen's proposal (Allen, 1995) is also in agreement with the demonstrated high light acclimatory roles of chloroplast two-component systems and their cyanobacterial homologues (Hsiao et al., 2004, Kappell and van Waasbergen, 2007).

Being part of the operon it controls is also a characteristic feature of the transcriptional feedback loop in circadian oscillators. We propose that the circadian analogy can be applied to chloroplasts, and that chloroplast encoded two component systems, acting as endogenous oscillators, generate a rhythmic pattern of chloroplast mRNA accumulation. A diurnal rhythm has been noted in photosynthetic electron transport (Okada and Horie, 1979). Chloroplast two-component systems, as discussed in the earlier section, connect the activity of photosynthetic electron transport chain to gene expression in chloroplasts. Thus it is likely that an autoregulatory loop in the transcription of chloroplast-encoded two-component systems, when connected to the rhythmic activity of photosynthetic electron transport, will generate a rhythmic transcript accumulation pattern for chloroplast genes.

An endogenous rhythm of chloroplast transcript accumulation has been demonstrated in the marine chromophytic alga *Heterosigma carterae* (Doran and Cattolico, 1997) and is thought to be generated by an endogenous oscillator driven by the activity of the photosynthetic electron transport chain. If our proposal is correct, chloroplasts in the green lineage may have lost this transcriptional loop as genes encoding chloroplast two-component systems have moved to the nuclear genome. The likely existence of a post-translational feedback loop in the nuclear-encoded chloroplast two-component systems may still ensure an endogenous rhythm of transcript accumulation in chloroplasts, as has been observed in *Arabidopsis* (Puthiyaveetil and Allen, 2008). Additionally, the observation that in CSK null mutants, rhythmcity of *psaA* transcript accumulation has been greatly attenuated

(Figure 5.1A) supports the role of nuclear-encoded chloroplast two-component system in rhythmic transcriptional activity of chloroplast genes.

The endogenous rhythm in chloroplasts is distinct from the circadian rhythm in that the former is aperiodic in oscillation. The aperiodicity in chloroplast's endogenous rhythm was demonstrated in algae, where the endogenous rhythm of transcript accumulation shows a temporally-gated response to changes in photoperiod (Doran and Cattolico, 1997). It was seen that the ability of the algae to crank up the transcriptional response, in response to changes in photoperiod, was limited to the first 2 hours of the dark period. Likewise, in the red alga, *Cyanidioschyzon merolae* a similar chloroplast transcriptional response was noted in response to shifts from dark to light conditions (Minoda et al., 2005). The transcription of photosystem genes and other chloroplast genes encoding core subunits of electron transfer complexes peaked after 1 hour from dark to light switch, which then decreased after 6 hours of the light shift. One can easily envisage how an endogenous oscillator made of chloroplast two-component systems can generate such aperiodic transcriptional rhythms in chloroplasts.

There could be clear advantages for the endogenous rhythm of transcript accumulation in chloroplasts not being circadian. For example, chloroplasts are faced with aperiodic fluctuations in incident light quantity and quality, such as those caused by cloud cover, transient shading by the canopy. In order to improve photosynthetic efficiency and to avoid free radical generation by inadvertent electron transfer reactions under such conditions, chloroplasts can rapidly tune the expression of genes encoding core subunits of the electron transfer complexes with the help of an aperiodic clock. As seen in the earlier section, chloroplast's ability to rapidly regulate the expression of chloroplast-encoded core subunits gives them the upper hand in the assembly of electron transfer complexes and hence in altering the thylakoid composition. Chloroplast's endogenous rhythm in transcript accumulation is thus driven by the availability of light and can be best described as a dial, or light-dark, cycle. It then appears that, besides providing the redox-signalling device that

connects photosynthesis with gene expression, two-component systems have endowed chloroplasts with an autonomous clock.

9.6 Future directions

9.6.1 Signal sensing and transduction mechanism of chloroplast sensor kinases

As discussed earlier, the N-terminal GAF and PAS domain respectively of CSK and ycf26 are presumably forming their redox sensor domains. As mentioned in chapter 5, GAF domains rely on certain redox-responsive cofactors for their redox sensing function. Interestingly, the secondary database, Prosite, predicts a nucleotide-binding domain called PUA within the GAF domain of the *Arabidopsis* CSK. This observation raises the possibility that the GAF domain in CSK bind one or more redox-responsive cofactors such as pyridine or flavin nucleotides for its redox sensing function. The redox sensing function of PAS domains has been attributed to redox-responsive cofactors such as haem (Gilles-Gonzalez and Gonzalez, 2004), flavin (Taylor, 2007) or conserved cysteine residues (Malpica et al., 2004). Whether the PAS domain of ycf26 houses any redox-responsive cofactors or contains redox-sensitive cysteine residues for its redox sensing function remains to be seen.

In order to investigate the redox sensing mechanism of CSK and ycf26, both *in vitro* and *in vivo* approaches should be explored. The *in vitro* approach may involve biochemical and spectroscopic studies with purified full-length CSK and ycf26 proteins and their isolated GAF and PAS domains. Redox-responsive prosthetic groups and cofactors, if present, can give unique colourations to the purified CSK and ycf26 proteins. Prosthetic groups and cofactors can also be identified by their spectral signatures. The redox state of possible prosthetic groups and cofactors can be modulated by the addition of specific redox agents and the resulting biochemical and structural changes in the GAF and PAS domain can be studied by visible, FTIR and NMR spectroscopy. The involvement of conserved cysteine residue in redox-sensing function can be studied by the use of redox and

alkylating agents. Structural elucidation of GAF and PAS domains can further give important insights into their redox sensing function.

The above *in vitro* approaches should be complemented by studies carried out *in vivo*. These would, however, require a model organism, which allows precise alterations of the *CSK* gene. The model moss *Physcomitrella patens* (Rensing et al., 2008) is an ideal system for *in vivo* studies as it permits gene targeting by homologous recombination. Homologous recombination in *Physcomitrella* enables rapid and efficient replacement of endogenous loci by precisely manipulated variant alleles differing by as little as a single base pair. This highly sophisticated gene replacement permits the analysis of variant genes placed under the control of the native promoter, *in situ*, within the genome, thereby avoiding potential complications resulting from "position effects" and epigenetic silencing. Additionally, *Physcomitrella* provides a phylogenetically distant model, affording an evolutionary perspective on the function of chloroplast two-component systems.

The critical amino acid residues of the sensor domain, once identified from biochemical and structural studies, should be altered by site-directed mutagenesis, and this is already possible in *Physcomitrella*. This approach would enable the study of the consequences of these alterations on the signalling role of CSK. Since ycf26 is predominantly found in non-green algae and since current genetic manipulation technology for these algae is unable to allow precise alterations in their genes, the characterization of ycf26 signalling role *in vivo* may have to wait. However, these studies should be possible to a certain degree in species such as diatoms, which shows some potential for genetic engineering.

Signal reception modulates the autokinase activity of sensor kinases. Modulation can be either activation or inactivation of the autokinase activity (Stock et al., 2000). For CSK and ycf26, it remains to be determined whether the signal perception causes activation or inactivation of their autophosphorylation activity. Autophosphorylation assays with purified, full-length proteins in the presence of

relevant signalling molecules can reveal whether the signal is an on or off switch for autophosphorylation. Having done this, the identity of the autophosphorylated amino acid in CSK and ycf26 should be determined by mass spectrometric analysis – LCMS – of the whole proteins and of peptide fragments.

The basis of signal transduction in two-component systems is phosphoryl group transfer from the sensor to the response regulator (Stock et al., 2000). The sensor kinase becomes dephosphorylated as its substrate, the phosphoryl group acceptor response regulator, becomes phosphorylated. Chloroplast sensor kinases should therefore be examined for the redox potential under which they become phosphorylated and dephosphorylated. In addition, the kinetics of phosphorylation and dephosphorylation should be examined. A labelled phosphorus atom transferred from one protein to another will show identical half-time of dephosphorylation in the donor and of phosphorylation in the acceptor. Over-expressed and purified chloroplast sensors and chloroplast response regulator proteins should therefore be used to study phosphoryl group transfer. The identity of the phosphorylated amino acids in chloroplast response regulators should also be determined by mass spectrometric analysis. The studies of autophosphorylation and phosphoryl group transfer acquire special significance in the case of CSK as it is a modified histidine kinase. Unconventional amino acids (discussed in chapter 6) and phosphoryl group transfer mechanism may be involved in the signal transduction of CSK. These studies in conjunction with structural determination of phosphorylated and unphosphorylated chloroplast sensors and response regulators can further reveal intricacies of their signalling mechanisms.

9.6.2 Identification of the cognate response regulator partner of CSK

As discussed in chapter 8, the cognate response regulator of CSK is only seen in red algae and diatoms, where CSK exists as an unmodified histidine kinase. When CSK becomes modified, as in green algae and plants, no cognate response regulators can yet be identified; however, CSK retains its transcriptional regulatory role.

Transcriptional regulation requires specific interaction with the transcriptional machinery and specialised proteins such as response regulators can only fulfil DNA binding properties. It can thus be concluded that a yet unidentified response regulator mediates the action of CSK in green algae and plants. The identification of this response regulator, however, is a challenging prospect.

Various methods can be employed for the identification of the cognate response regulator partner of CSK in green algae and plants. Yeast two-hybrid screening and split ubiquitin screening are methods that are routinely used for the identification of interaction partners. These methods, however, can present some problems. Since both the "bait" and "prey" proteins are overexpressed, non-specific interaction may result from their overabundance in the yeast or bacterial cell. Additionally, the presence of unprocessed transit peptides in the bait and prey proteins can inhibit their intermolecular interaction in yeast cells.

Candidate chloroplast response regulators with chloroplast targeting sequence can also be identified by bioinformatic analysis. The possibility that the response regulator-like protein TCP34 functions as the cognate partner of CSK should be examined by inactivation of the *TCP34* gene in both *Arabidopsis* and *Physcomitrella*. If the functional studies of TCP34 null mutants suggest TCP34 to be a genuine chloroplast response regulator, the interaction between TCP34 and CSK should be studied in *Physcomitrella* by dual epitope-tagging and co-immunoprecipitation. TCP34 can be tagged with one epitope and CSK with another. The co-immunoprecipitation of these two separate epitopes should reveal whether TCP34 and CSK interact. Additionally, the interaction between CSK and TCP34 can be validated *in vivo* by split YFP strategy.

If it turns out that chloroplast response regulators are indeed lost in green algae and plants, how would the regulatory action of CSK be relayed to the transcriptional machinery in their chloroplasts? In other words, what will be the effector component that translates CSK's sensing function into transcriptional

activation/repression of chloroplast genes? The possibility that CSK directly phosphorylates or interacts with the transcriptional machinery cannot be ruled out. Under such circumstances, the subunits of the plastid encoded polymerase (PEP) or its cofactors such as sigma-like factors could act as phosphorylation targets of CSK. It has been known that the redox-responsive, PEP-associated serine/threonine kinase, plastid transcription kinase (PEP) acts via phosphorylation of sigma-like factors and of subunits of the PEP (Baginsky et al., 1997).

9.6.3 Determination of the target genes and mode of action of chloroplast response regulators

Chloroplast genomes encode as many as 200 genes in certain algal species or as little as a few dozen in some non-photosynthetic organisms. Land plant chloroplast genomes typically contain around 100-120 genes. It is thus interesting to ask how many of these genes are under the direct control of chloroplast response regulators? Since not all chloroplast genes may be transcriptionally regulated under a given condition, different experimental conditions should be used to assess the regulation of chloroplast gene expression. In this regard, use of photosystem-specific lights (light 1 and 2), changes in light fluence rate (light quantity), electron transport inhibitors, and various stages in a photoperiod are relevant conditions to assess the chloroplast transcriptome. A chloroplast DNA microarray analysis is an ideal way to find out the number of chloroplast genes under the regulatory control of chloroplast response regulators. Additionally, qPCR, northern analysis and run-on transcriptional assays can be used to validate transcriptional control of selected chloroplast genes.

Since response regulators affect transcription by way of selective DNA binding or by interacting with RNA polymerases, the DNA binding properties of chloroplast response regulators should be studied by electrophoretic mobility shift assays and DNase footprinting assays. The plastid array data will also reveal common *cis*-acting elements in the target genes of chloroplast response regulators. It has been recognized that response regulators exist in two conformations, an active

and an inactive conformation (Robinson et al., 2000). Phosphorylation in the receiver domain of response regulators relieves the inhibitory effect on their DNA-binding domains imposed by the receiver and thereby shifts the equilibrium of response regulator conformation into that of an active state. The response regulators, in addition to catalysing the phosphate group transfer from phosphohistidine of sensor kinases, catalyse the hydrolysis of their own phosphate group. The lifetime of the acyl phosphate in response regulators is known to vary from seconds to hours depending on the specific needs of each signalling pathway (Robinson et al., 2000). It will therefore be interesting to determine the lifetime of acyl phosphates in chloroplast response regulators and what this lifetime means for the regulation of chloroplast transcription. Some response regulators are known to act as activators as well as repressors of transcription by binding at more than one promoter sites or with varying affinity (Lynch and Lin, 1996, Li et al., 2008). This dual action is an interesting prospect for chloroplast response regulators since only a bifunctional chloroplast response regulator can account for both upregulation and downregulation seen in some light conditions (Pfannschmidt et al., 1999a). A complete description of the action of chloroplast response regulators will also require determination of their three-dimensional structures.

9.6.4 Development of a reporter gene based functional assay for chloroplast two-component systems

Development of a chloroplast reporter gene-based assay is an important avenue to be explored for understanding the full implications of chloroplast two-component systems for regulating chloroplast transcription. A chloroplast reporter gene-based assay can reveal hitherto inaccessible details in the function of chloroplast two-component systems. This is in addition to the fact that the present analysis of the functional role of chloroplast two-component systems is hampered by inconsistencies and difficulties associated with the determination of chloroplast transcript abundance and transcriptional activity. A chloroplast reporter-gene based assay would also enable monitoring of chloroplast transcription in real time.

Various reporter genes have been assayed in chloroplast after their successful integration and expression under native chloroplast or foreign promoters. Both luciferase and GFP reporter genes have been assayed in the chloroplast of *Chlamydomonas* (Mayfield and Schultz, 2004, Franklin et al., 2002). Biolistic transformation of tobacco chloroplasts with the reporter gene *UidA*, which codes for the beta glucuronidase enzyme used in GUS assay, has been achieved (Bock and Maliga, 1995). GFP reporter gene has also been successfully delivered into tobacco chloroplasts with gene bombardment technology (Newell et al., 2003). However, chloroplast transformation technologies have not been very successful for model plants such as *Arabidopsis*. The efficiency of biolistic transformation was found to be very low in *Arabidopsis* (Sikdar et al., 1998). Generating homoplasmic plants for the analysis would also require tissue culture regeneration of whole, fertile plants from single transformant cells, which has proved difficult for *Arabidopsis* until now (Sikdar et al., 1998).

Combining the success of plastid transformation with the availability of knockout mutants for two-component systems in a single experimental system is thus the stumbling block in a chloroplast reporter gene-based assay for chloroplast two-component systems. However, this is going to change soon as experimental organisms such as *Physcomitrella* are amenable to both gene inactivation and plastid transformation (Cho et al., 1999). If and when such a system is available, the consequences of the inactivation of chloroplast two-component systems on chloroplast transcription can be studied by fusing a reporter gene with photosystem gene promoters such as that of *psaA*, which responds well to the changes in the redox state of PQ. Reporter genes can also be expressed under gene promoters of *rbcL*, *psbA* or *rrn16*.

Genotyping CSK T-DNA mutants for homozygous insertion lines

The Phenotypes identified and described for Chloroplast Sensor Kinase (CSK) were based on two completely characterized T-DNA lines, SALK_027360 and SALK_018074, which harbor insertions in the *At1g67840* locus encoding the CSK protein. Seeds for SALK_027360 and SALK_018074 lines were obtained from the European Arabidopsis Stock Center (Nottingham), sown on soil and the F_1 plants were selfed. A PCR based approach (McKinney et al., 1995) was used to identify homozygous insertion lines amongst the F_1 plants. Seeds obtained from the genotyped, F_1 homozygous plants were used to characterize the phenotype. The PCR based genotyping approach involved a genomic PCR using genomic primers, a second PCR using genomic and T-DNA left border primers and a third Reverse Transcriptase PCR (RT-PCR) to confirm the absence of transcripts from the *At1g67840* locus.

The positions of the T-DNA insertion in SALK_027360 and SALK_018074 lines are indicated in Figure 1. Figure 1B shows results from a genomic PCR using 5' tggcctcttttagctatgggga 3' as the forward primer and 5' tgctcaagacaaagccgttga 3' as the reverse primer. The wild type *CSK* gene is amplified from the wild type sample and not from the SALK_027360 sample indicating T-DNA insertion. Figure 1C shows results from a genomic PCR using 5' tggcctcttttagctatgggga 3' as forward primer and 5' gcgtggaccgcttgctgcaact 3' as reverse primer, which is also the T-DNA cassette left border primer. The length of the amplificon from this second PCR reaction is indicative of the approximate insertion site in the SALK_027360. The actual insertion site was further determined by sequencing the amplificon from this PCR reaction. Figure 1D shows results from a reverse transcriptase PCR using 5' gagagtttcagtctcagccaca 3' as forward primer and 5' ttgcaatcaattttgttcaagtc 3'as reverse primer. The *CSK* mRNA is amplified only from the wild type and not from the SALK_027360 line, thus confirming that the *At1g67840* locus is not transcribed in SALK_027360 line.

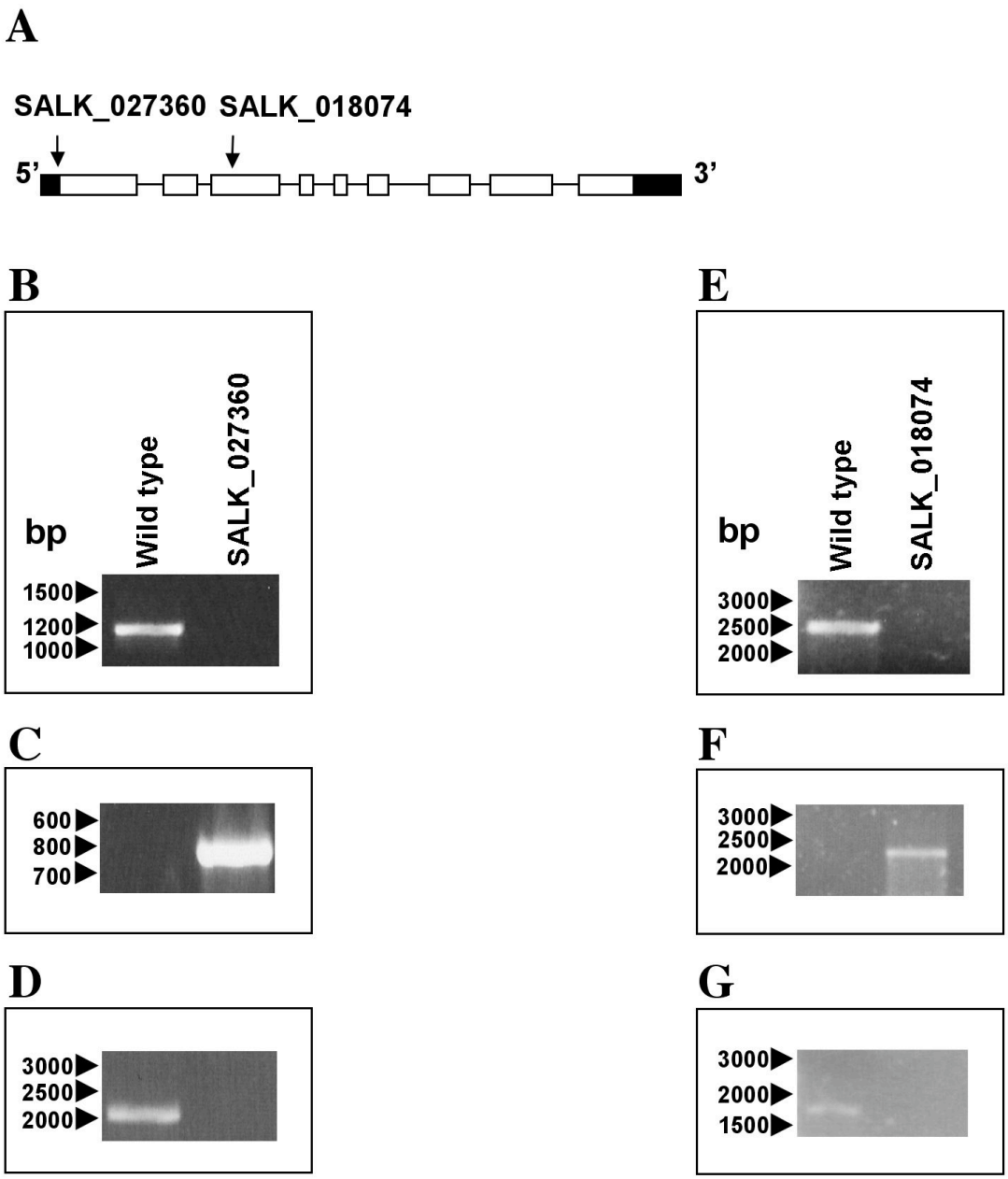


Figure 1 Genotyping of CSK T-DNA lines. (**A**) Schematic representation of the gene region of *CSK* indicating T-DNA insertion sites in two different SALK lines. Exons are represented as white rectangles and introns as the lines connecting them. The UTR regions at either end of the transcript are shown as filled rectangles. (**B** to **G**) Results from the genomic and the Reverse Transcriptase PCR methods used for genotyping SALK_027360 and SALK_018074 lines. Reactions products were separated on a 1% agarose gel and stained with EtBr. Position of DNA ladder (in base pair, bp) is indicated on the left.

Figure 1E shows results from a genomic PCR using 5' gtagagttcacacagatgattgagaaa 3' as the forward primer and 5' gcttcattggcttcagatactgc 3' as the reverse primer. The wild type *CSK* gene is amplified from the wild type sample and not from the SALK_018074 sample, indicating T-DNA insertion. Figure 1F shows results from a genomic PCR using 5' gcgtggaccgcttgctgcaact 3' as the forward primer, which is also the T-DNA cassette left border primer and 5' gcttcattggcttcagatactgc 3' as the reverse primer. The length of the amplificon from this PCR reaction is indicative of the approximate insertion site in the SALK_018074. The actual insertion site was further determined by sequencing the amplificon from this PCR reaction. Figure 1G shows results from a reverse transcriptase PCR using 5' atgcttctttctgcaatcgc 3' as the forward primer and 5' ctatgcttcattggctttcag 3'as the reverse primer. The *CSK* mRNA became amplified only from the wild type and not from the SALK_018074 line, thus confirming that the *At1g67840* locus is not transcribed in the SALK_018074 line.

ABRAHAMS, J. P., LESLIE, A. G., LUTTER, R. & WALKER, J. E. (1994) Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature,* 370**,** 621-8.

ALLEN, J. F. (1993a) Control of Gene-Expression by Redox Potential and the Requirement for Chloroplast and Mitochondrial Genomes. *J. Theor. Biol.,* 165**,** 609-631.

ALLEN, J. F. (1993b) Redox Control of Transcription - Sensors, Response Regulators, Activators and Repressors. *FEBS Lett.,* 332**,** 203-207.

ALLEN, J. F. (1995) Thylakoid Protein-Phosphorylation, State-1-State-2 Transitions, and Photosystem Stoichiometry Adjustment - Redox Control at Multiple Levels of Gene-Expression. *Physiologia Plantarum,* 93**,** 196-205.

ALLEN, J. F. (2003a) Botany. State transitions--a question of balance. *Science,* 299**,** 1530-2.

ALLEN, J. F. (2003b) The function of genomes in bioenergetic organelles. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.,* 358**,** 19-37.

ALLEN, J. F. (2004) Cytochrome b6f: structure for signalling and vectorial metabolism. *Trends Plant Sci,* 9**,** 130-7.

ALLEN, J. F., BENNETT, J., STEINBACK, K. E. & ARNTZEN, C. J. (1981) Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems. *Nature,* 291**,** 25-29.

ALLEN, J. F. & FORSBERG, J. (2001) Molecular recognition in thylakoid structure and function. *Trends Plant Sci,* 6**,** 317-26.

ALLEN, J. F. & MARTIN, W. (2007) Evolutionary biology: out of thin air. *Nature,* 445**,** 610-2.

ALLEN, J. F. & NILSSON, A. (1997) Redox signalling and the structural basis of regulation of photosynthesis by protein phosphorylation. *Physiologia Plantarum,* 100**,** 863-868.

ALLEN, J. F. & RAVEN, J. A. (1996) Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J Mol Evol,* 42**,** 482-92.

ALONSO, J. M. & STEPANOVA, A. N. (2004) The ethylene signaling pathway. *Science,* 306**,** 1513-5.

ALONSO, J. M., STEPANOVA, A. N., LEISSE, T. J., KIM, C. J., CHEN, H., SHINN, P., STEVENSON, D. K., ZIMMERMAN, J., BARAJAS, P., CHEUK, R., GADRINAB, C., HELLER, C., JESKE, A., KOESEMA, E., MEYERS, C. C., PARKER, H., PREDNIS, L., ANSARI, Y., CHOY, N., DEEN, H., GERALT, M.,

HAZARI, N., HOM, E., KARNES, M., MULHOLLAND, C., NDUBAKU, R., SCHMIDT, I., GUZMAN, P., AGUILAR-HENONIN, L., SCHMID, M., WEIGEL, D., CARTER, D. E., MARCHAND, T., RISSEEUW, E., BROGDEN, D., ZEKO, A., CROSBY, W. L., BERRY, C. C. & ECKER, J. R. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science,* 301**,** 653-7.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. (1990) Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

AMUNTS, A., DRORY, O. & NELSON, N. (2007) The structure of a plant photosystem I supercomplex at 3.4 A resolution. *Nature,* 447**,** 58-63.

ANDERSON, C. (1992) Nobel Prize given for work on protein phosphorylation. *Nature,* 359**,** 570.

ARAVIND, L. & PONTING, C. P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci,* 22**,** 458-9.

ARAVIND, L. & PONTING, C. P. (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol Lett,* 176**,** 111-6.

ARMBRUST, E. V., BERGES, J. A., 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., KROGER, N., LAU, W. W., LANE, T. W., LARIMER, F. W., LIPPMEIER, J. C., LUCAS, S., MEDINA, M., MONTSANT, A., OBORNIK, M., PARKER, M. S., 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.

ASHBY, M., HOUMARD, J. & MULLINEAUX, C. (2002) The ycf27 genes from cyanobacteria and eukaryotic algae: distribution and implications for chloroplast evolution. *FEMS Microbiol. Lett.,* 214**,** 25.

ASHBY, M. K. & HOUMARD, J. (2006) Cyanobacterial two-component proteins: structure, diversity, distribution, and evolution. *Microbiol. Mol. Biol. Rev.,* 70**,** 472-509.

ASHBY, M. K. & MULLINEAUX, C. W. (1999) Cyanobacterial ycf27 gene products regulate energy transfer from phycobilisomes to photosystems I and II. *FEMS Microbiol Lett,* 181**,** 253-60.

BAENA-GONZALEZ, E., BAGINSKY, S., MULO, P., SUMMER, H., ARO, E. M. & LINK, G. (2001) Chloroplast transcription at different light intensities. Glutathione-mediated phosphorylation of the major RNA polymerase involved in redox-regulated organellar gene expression. *Plant Physiol,* 127**,** 1044-52.

BAGINSKY, S., TILLER, K. & LINK, G. (1997) Transcription factor phosphorylation by a protein kinase associated with chloroplast RNA polymerase from mustard (Sinapis alba). *Plant Mol Biol,* 34**,** 181-9.

BANNAI, H., TAMADA, Y., MARUYAMA, O., NAKAI, K. & MIYANO, S. (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics,* 18**,** 298-305.

BAUER, C., ELSEN, S., SWEM, L. R., SWEM, D. L. & MASUDA, S. (2003) Redox And Light Regulation Of Gene Expression In Photosynthetic Prokaryotes. *Phil. Trans. R. Soc. Lond. B.*

BAUER, C. E., ELSEN, S. & BIRD, T. H. (1999) Mechanisms for redox control of gene expression. *Annu. Rev. Microbiol.,* 53**,** 495-523.

BEDBROOK, J. R. & BOGORAD, L. (1976) Endonuclease recognition sites mapped on Zea mays chloroplast DNA. *Proc Natl Acad Sci U S A,* 73**,** 4309-4313.

BEKKER, M., TEIXEIRA DE MATTOS, M. J. & HELLINGWERF, K. J. (2006) The role of two-component regulation systems in the physiology of the bacterial cell. *Sci Prog,* 89**,** 213-42.

BERGERAT, A., DE MASSY, B., GADELLE, D., VAROUTAS, P. C., NICOLAS, A. & FORTERRE, P. (1997) An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature,* 386**,** 414-7.

BILWES, A. M., ALEX, L. A., CRANE, B. R. & SIMON, M. I. (1999) Structure of CheA, a signal-transducing histidine kinase. *Cell,* 96**,** 131-41.

BOCK, A. & GROSS, R. (2002) The unorthodox histidine kinases BvgS and EvgS are responsive to the oxidation status of a quinone electron carrier. *Eur J Biochem,* 269**,** 3479-84.

BOCK, R. & MALIGA, P. (1995) Correct splicing of a group II intron from a chimeric reporter gene transcript in tobacco plastids. *Nucleic Acids Res,* 23**,** 2544-7.

BOGORAD, L. (1998) Discovery of Chloroplast DNA, Genomes and Genes. IN KUNG, S.-D. & YANG, S.-F. (Eds.) *Discoveries in Plant Biology*. Singapore, World Scientific.

BOUDREAU, E., TAKAHASHI, Y., LEMIEUX, C., TURMEL, M. & ROCHAIX, J. D. (1997) The chloroplast ycf3 and ycf4 open reading frames of Chlamydomonas reinhardtii are required for the accumulation of the photosystem I complex. *EMBO J,* 16**,** 6095-104.

BROCK, I. W., HAZELL, L., MICHL, D., NIELSEN, V. S., MOLLER, B. L., HERRMANN, R. G., KLOSGEN, R. B. & ROBINSON, C. (1993) Precursors of one integral and five lumenal thylakoid proteins are imported by isolated pea and barley thylakoids: optimisation of in vitro assays. *Plant Mol Biol,* 23**,** 717-25.

BRUCE, B. D. (2001) The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim Biophys Acta,* 1541**,** 2-21.

BUCHANAN, B. (1980) Role of light in the regulation of chloroplast enzymes *Annu. Rev. Plant Physiol. ,* 31**,** 341-374.

BUSS, K. A., COOPER, D. R., INGRAM-SMITH, C., FERRY, J. G., SANDERS, D. A. & HASSON, M. S. (2001) Urkinase: structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases. *J Bacteriol,* 183**,** 680-6.

CANN, M. (2007) A subset of GAF domains are evolutionarily conserved sodium sensors. *Mol Microbiol,* 64**,** 461-72.

CATLETT, N. L., YODER, O. C. & TURGEON, B. G. (2003) Whole-genome analysis of two-component signal transduction genes in fungal pathogens. *Eukaryot Cell,* 2**,** 1151-61.

CHANG, C., KWOK, S. F., BLEECKER, A. B. & MEYEROWITZ, E. M. (1993) Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. *Science,* 262**,** 539-44.

CHANG, C. & STEWART, R. C. (1998) The two-component system. Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiol,* 117**,** 723-31.

CHO, S. H., CHUNG, Y. S., CHO, S. K., RIM, Y. W. & SHIN, J. S. (1999) Particle bombardment mediated transformation and GFP expression in the moss Physcomitrella patens. *Mol Cells,* 9**,** 14-9.

CHOQUET, Y. & VALLON, O. (2000) Synthesis, assembly and degradation of thylakoid membrane proteins. *Biochimie,* 82**,** 615-34.

CHOW, W. S., MELIS, A. & ANDERSON, J. M. (1990) Adjustments of Photosystem Stoichiometry in Chloroplasts Improve the Quantum Efficiency of Photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 7502-7506.

CIKOS, S., BUKOVSKA, A. & KOPPEL, J. (2007) Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis. *BMC Mol Biol,* 8**,** 113.

CLAMP, M., CUFF, J., SEARLE, S. M. & BARTON, G. J. (2004) The Jalview Java alignment editor. *Bioinformatics,* 20**,** 426-7.

COLE, C., BARBER, J. D. & BARTON, G. J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*.

COLEMAN, J. E. (1992) Structure and mechanism of alkaline phosphatase. *Annu Rev Biophys Biomol Struct,* 21**,** 441-83.

D'AUTREAUX, B., TUCKER, N. P., DIXON, R. & SPIRO, S. (2005) A non-haem iron centre in the transcription factor NorR senses nitric oxide. *Nature,* 437**,** 769-72.

DANON, A. & MAYFIELD, S. P. (1994) Light-regulated translation of chloroplast messenger RNAs through redox potential. *Science,* 266**,** 1717-9.

DEININGER, W., KROGER, P., HEGEMANN, U., LOTTSPEICH, F. & HEGEMANN, P. (1995) Chlamyrhodopsin represents a new type of sensory photoreceptor. *EMBO J,* 14**,** 5849-58.

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. U. S. A.,* 103**,** 11647-52.

DI COLA, A., KLOSTERMANN, E. & ROBINSON, C. (2005) The complexity of pathways for protein import into thylakoids: it's not easy being green. *Biochem. Soc. Trans.,* 33**,** 1024-7.

DORAN, E. & CATTOLICO, R. A. (1997) Photoregulation of Chloroplast Gene Transcription in the Chromophytic Alga Heterosigma carterae. *Plant Physiol,* 115**,** 773-781.

DUFRESNE, A., SALANOUBAT, M., PARTENSKY, F., ARTIGUENAVE, F., AXMANN, I. M., BARBE, V., DUPRAT, S., GALPERIN, M. Y., KOONIN, E. V., LE GALL, F., MAKAROVA, K. S., OSTROWSKI, M., OZTAS, S., ROBERT, C., ROGOZIN, I. B., SCANLAN, D. J., TANDEAU DE MARSAC, N., WEISSENBACH, J., WINCKER, P., WOLF, Y. I. & HESS, W. R. (2003) Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci. U. S. A.,* 100**,** 10020-5.

DUPLESSIS, M. R., KAROL, K. G., ADMAN, E. T., CHOI, L. Y., JACOBS, M. A. & CATTOLICO, R. A. (2007) Chloroplast His-to-Asp signal transduction: a potential mechanism for plastid gene regulation in Heterosigma akashiwo (Raphidophyceae). *BMC Evol. Biol.,* 7**,** 70.

DUTTA, R. & INOUYE, M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci,* 25**,** 24-8.

DUVE, C. D. (1991) *Blueprint for a Cell: The Nature and Origin of Life,* Burlington, N.C., Neil Patterson.

EL BISSATI, K. & KIRILOVSKY, D. (2001) Regulation of psbA and psaE expression by light quality in Synechocystis species PCC 6803. A redox control mechanism. *Plant Physiology,* 125**,** 1988-2000.

ELLIS, R. J. (1984) The Nuclear Domination of Chloroplast Development. *Sci. Prog. Oxf.,* 69**,** 129-142.

EMANUELSSON, O., NIELSEN, H., BRUNAK, S. & VON HEIJNE, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol,* 300**,** 1005-16.

EMANUELSSON, O., NIELSEN, H. & VON HEIJNE, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci,* 8**,** 978-84.

ESCOUBAS, J. M., LOMAS, M., LAROCHE, J. & FALKOWSKI, P. G. (1995) Light-Intensity Regulation of Cab Gene-Transcription Is Signaled by the Redox State of the Plastoquinone Pool. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 10237-10241.

FERREIRA, K. N., IVERSON, T. M., MAGHLAOUI, K., BARBER, J. & IWATA, S. (2004) Architecture of the photosynthetic oxygen-evolving center. *Science,* 303**,** 1831-8.

FEY, V., WAGNER, R., BRAUTIGAM, K. & PFANNSCHMIDT, T. (2005a) Photosynthetic redox control of nuclear gene expression. *J Exp Bot,* 56**,** 1491-8.

FEY, V., WAGNER, R., BRAUTIGAM, K., WIRTZ, M., HELL, R., DIETZMANN, A., LEISTER, D., OELMULLER, R. & PFANNSCHMIDT, T. (2005b) Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of Arabidopsis thaliana. *J Biol Chem,* 280**,** 5318-28.

FINN, R. D., MISTRY, J., SCHUSTER-BOCKLER, B., GRIFFITHS-JONES, S., HOLLICH, V., LASSMANN, T., MOXON, S., MARSHALL, M., KHANNA, A., DURBIN, R., EDDY, S. R., SONNHAMMER, E. L. & BATEMAN, A. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res,* 34**,** D247-51.

FORSBERG, J., ROSENQUIST, M., FRAYSSE, L. & ALLEN, J. F. (2001) Redox signalling in chloroplasts and mitochondria: genomic and biochemical evidence for two-component regulatory systems in bioenergetic organelles. *Biochem Soc Trans,* 29**,** 403-7.

FRANKLIN, S., NGO, B., EFUET, E. & MAYFIELD, S. P. (2002) Development of a GFP reporter gene for Chlamydomonas reinhardtii chloroplast. *Plant J,* 30**,** 733-44.

FRIEDMAN, A. L. & KEEGSTRA, K. (1989) Chloroplast Protein Import - Quantitative-Analysis of Precursor Binding. *Plant Physiology,* 89**,** 993-999.

FUHRMANN, M., STAHLBERG, A., GOVORUNOVA, E., RANK, S. & HEGEMANN, P. (2001) The abundant retinal protein of the Chlamydomonas eye is not the photoreceptor for phototaxis and photophobic responses. *J Cell Sci,* 114**,** 3857-63.

FUJITA, Y. (1997) A study on the dynamic features of photosystem stoichiometry: Accomplishments and problems for future studies. *Photosynthesis Research,* 53**,** 83-93.

GEORGELLIS, D., KWON, O. & LIN, E. C. (2001) Quinones as the redox signal for the arc two-component system of bacteria. *Science,* 292**,** 2314-6.

GIBBS, S. P. (1981) The Chloroplasts of Some Algal Groups May Have Evolved from Endosymbiotic Eukaryotic Algae. *Annal. N. Y. Acad. Sci.,* 361**,** 193-208.

GILLES-GONZALEZ, M. A. & GONZALEZ, G. (2004) Signal transduction by heme-containing PAS-domain proteins. *J Appl Physiol,* 96**,** 774-83.

GLAZER, A. N. (1985) Light harvesting by phycobilisomes. *Annu Rev Biophys Biophys Chem,* 14**,** 47-77.

GOLDSCHMIDTCLERMONT, M. (1998) Coordination of nuclear and chloroplast gene expression in plant cells. *International Review of Cytology - a Survey of Cell Biology*.

GOULD, S. B., WALLER, R. F. & MCFADDEN, G. I. (2008) Plastid evolution. *Annu Rev Plant Biol,* 59**,** 491-517.

GOVINDJEE (2000) Milestones in Photosynthesis Research. IN YUNUS, M., PATHRE, U. & MOHANTY, P. (Eds.) *Probing Photosynthesis: Mechanisms, Regulation And Adaptation*. London, Taylor & Francis.

GRABBE, R. & SCHMITZ, R. A. (2003) Oxygen control of nif gene expression in Klebsiella pneumoniae depends on NifL reduction at the cytoplasmic membrane by electrons derived from the reduced quinone pool. *Eur J Biochem,* 270**,** 1555-66.

GRAY, J. C., SULLIVAN, J. A., WANG, J. H., JEROME, C. A. & MACLEAN, D. (2003) Coordination of plastid and nuclear gene expression. *Philos Trans R Soc Lond B Biol Sci,* 358**,** 135-44; discussion 144-5.

GREEN, B. R., PICHERSKY, E. & KLOPPSTECH, K. (1991) Chlorophyll a/b-binding proteins: an extended family. *Trends Biochem Sci,* 16**,** 181-6.

GREEN, J. & PAGET, M. S. (2004) Bacterial redox sensors. *Nat Rev Microbiol,* 2**,** 954-66.

GRUISSEM, W. & TONKYN, J. C. (1993) Control Mechanisms of Plastid Gene-Expression. *Critical Reviews in Plant Sciences,* 12**,** 19-55.

GUINDON, S. & GASCUEL, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology,* 52**,** 696-704.

HANKAMER, B., MORRIS, E. P. & BARBER, J. (1999) Revealing the structure of the oxygen-evolving core dimer of photosystem II by cryoelectron crystallography. *Nat Struct Biol,* 6**,** 560-4.

HELLIWELL, C. A., SULLIVAN, J. A., MOULD, R. M., GRAY, J. C., PEACOCK, W. J. & DENNIS, E. S. (2001) A plastid envelope location of Arabidopsis ent-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway. *The Plant Journal,* 28**,** 201-208.

HERRANEN, M., TYYSTJARVI, T. & ARO, E. M. (2005) Regulation of photosystem I reaction center genes in Synechocystis sp. strain PCC 6803 during Light acclimation. *Plant Cell Physiol,* 46**,** 1484-93.

HIBBERD, J. M., LINLEY, P. J., KHAN, M. S. & GRAY, J. C. (1998) Transient expression of green fluorescent protein in various plastid types following microprojectile bombardment. *The Plant Journal,* 16**,** 627–632.

HIHARA, Y., SONOIKE, K., KANEHISA, M. & IKEUCHI, M. (2002) Search for redox responsive genes in Synechocystis sp. PCC 6803 by using DNA microarrays. *Plant and Cell Physiology,* 43**,** S178-S178.

HILL, R. & BENDALL, F. (1960) Function of the two cytochrome components in chloroplasts - a working hypothesis. *Nature,* 186**,** 136-137.

HO, Y. S., BURDEN, L. M. & HURLEY, J. H. (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *Embo J,* 19**,** 5288-99.

HOCH, J. A. (2000) Two-component and phosphorelay signal transduction. *Curr Opin Microbiol,* 3**,** 165-70.

HORTON, P., PARK, K. J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C. J. & NAKAI, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res,* 35**,** W585-7.

HSIAO, H. Y., HE, Q., VAN WAASBERGEN, L. G. & GROSSMAN, A. R. (2004) Control of photosynthetic and high-light-responsive genes by the histidine kinase DspA: negative and positive regulation and interactions between signal transduction pathways. *J Bacteriol,* 186**,** 3882-8.

HUANG, C. Y., AYLIFFE, M. A. & TIMMIS, J. N. (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature,* 422**,** 72-6.

HWANG, I., CHEN, H. C. & SHEEN, J. (2002) Two-component signal transduction pathways in Arabidopsis. *Plant Physiol,* 129**,** 500-15.

IUCHI, S. & LIN, E. C. C. (1992) Mutational Analysis of Signal Transduction by Arcb, a Membrane Sensor Protein Responsible for Anaerobic Repression of Operons Involved in the Central Aerobic Pathways in Escherichia-Coli. *J. Bacteriol.,* 174**,** 3972-3980.

IVLEVA, N. B., GAO, T., LIWANG, A. C. & GOLDEN, S. S. (2006) Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proc Natl Acad Sci U S A,* 103**,** 17468-73.

JACOBS, M. A., CONNELL, L. & CATTOLICO, R. A. (1999) A conserved His-Asp signal response regulator-like gene in Heterosigma akashiwo chloroplasts. *Plant Mol Biol,* 41**,** 645-55.

JAMES, H. E., BARTLING, D., MUSGROVE, J. E., KIRWIN, P. M., HERRMANN, R. G. & ROBINSON, C. (1989) Transport of proteins into chloroplasts. Import and maturation of precursors to the 33-, 23-, and 16-kDa proteins of the photosynthetic oxygen-evolving complex. *J Biol Chem,* 264**,** 19573-6.

JANSSON, S. (1999) A guide to the Lhc genes and their relatives in Arabidopsis/IT>. *Trends Plant Sci,* 4**,** 236-240.

JUNGE, W., LILL, H. & ENGELBRECHT, S. (1997) ATP synthase: an electrochemical transducer with rotatory mechanics. *Trends Biochem Sci,* 22**,** 420-3.

KAPPELL, A. D. & VAN WAASBERGEN, L. G. (2007) The response regulator RpaB binds the high light regulatory 1 sequence upstream of the high-light-inducible hliB gene from the cyanobacterium Synechocystis PCC 6803. *Arch Microbiol,* 187**,** 337-42.

KESSLER, U., MAID, U. & ZETSCHE, K. (1992) An equivalent to bacterial ompR genes is encoded on the plastid genome of red algae. *Plant Mol Biol,* 18**,** 777-80.

KIM, M., THUM, K. E., MORISHIGE, D. T. & MULLET, J. E. (1999) Detailed architecture of the barley chloroplast psbD-psbC blue light-responsive promoter. *J Biol Chem,* 274**,** 4684-92.

KIRK, J. T. O. & TILNEY-BASSETT, R. A. E. (1978) *The plastids. Their chemistry, structure, growth and inheritance,* Amsterdam, Elsevier/North Holland.

KLEFFMANN, T., HIRSCH-HOFFMANN, M., GRUISSEM, W. & BAGINSKY, S. (2006) plprot: a comprehensive proteome database for different plastid types. *Plant Cell Physiol,* 47**,** 432-6.

KLUMPP, S. & KRIEGLSTEIN, J. (2002) Phosphorylation and dephosphorylation of histidine residues in proteins. *Eur J Biochem,* 269**,** 1067-71.

KORETKE, K. K., LUPAS, A. N., WARREN, P. V., ROSENBERG, M. & BROWN, J. R. (2000) Evolution of two-component signal transduction. *Mol Biol Evol,* 17**,** 1956-70.

KOUSSEVITZKY, S., NOTT, A., MOCKLER, T. C., HONG, F., SACHETTO-MARTINS, G., SURPIN, M., LIM, J., MITTLER, R. & CHORY, J. (2007) Signals from chloroplasts converge to regulate nuclear gene expression. *Science,* 316**,** 715-9.

KOVACS, L., WIESSNER, W., KIS, M., NAGY, F., MENDE, D. & DEMETER, S. (2000) Short- and long-term redox regulation of photosynthetic light energy distribution and photosystem stoichiometry by acetate metabolism in the green alga, Chlamydobotrys stellata. *Photosynthesis Research,* 65**,** 231-247.

KUMAR, A., TOLEDO, J. C., PATEL, R. P., LANCASTER, J. R., JR. & STEYN, A. J. (2007) Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor. *Proc Natl Acad Sci U S A,* 104**,** 11568-73.

KURISU, G., ZHANG, H., SMITH, J. L. & CRAMER, W. A. (2003) Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science,* 302**,** 1009-14.

LANE, N. (2002) *Oxygen,* Oxford, Oxford University Press.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-8.

LASKER, M. V., THAI, P., BESANT, P. G., BUI, C. D., NAIDU, S. & TURCK, C. W. (2002) Branched-Chain _-Ketoacid Dehydrogenase Kinase: A Mammalian Enzyme With Histidine Kinase Activity. *J Biomol Tech,* 13**,** 238-245.

LI, H. & SHERMAN, L. A. (2000) A redox-responsive regulator of photosynthesis gene expression in the cyanobacterium Synechocystis sp. Strain PCC 6803. *J. Bacteriol.,* 182**,** 4268-4277.

LI, L., ALVEY, R. M., BEZY, R. P. & KEHOE, D. M. (2008) Inverse transcriptional activities during complementary chromatic adaptation are controlled by the response regulator RcaC binding to red and green light-responsive promoters. *Mol Microbiol,* 68**,** 286-97.

LITTLE, R. & DIXON, R. (2003) The amino-terminal GAF domain of Azotobacter vinelandii NifA binds 2-oxoglutarate to resist inhibition by NifL under nitrogen-limiting conditions. *J Biol Chem,* 278**,** 28711-8.

LOPEZ-JUEZ, E. (2007) Plastid biogenesis, between light and shadows. *J Exp Bot,* 58**,** 11-26.

LOPEZ-JUEZ, E., BOWYER, J. R. & SAKAI, T. (2007) Distinct leaf developmental and gene expression responses to light quantity depend on blue-photoreceptor or plastid-derived signals, and can occur in the absence of phototropins. *Planta,* 227**,** 113-23.

LOPEZ-JUEZ, E. & PYKE, K. A. (2005) Plastids unleashed: their development and their integration in plant development. *Int J Dev Biol,* 49**,** 557-77.

LOPEZ-MAURY, L., GARCIA-DOMINGUEZ, M., FLORENCIO, F. J. & REYES, J. C. (2002) A two-component signal transduction system involved in nickel sensing in the cyanobacterium Synechocystis sp. PCC 6803. *Mol Microbiol,* 43**,** 247-56.

LUNDE, C., JENSEN, P. E., HALDRUP, A., KNOETZEL, J. & SCHELLER, H. V. (2000) The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. *Nature,* 408**,** 613-5.

LYNCH, A. S. & LIN, E. C. (1996) Transcriptional control mediated by the ArcA two-component response regulator protein of Escherichia coli: characterization of DNA binding at target promoters. *J Bacteriol,* 178**,** 6238-49.

LYSENKO, E. A. & KUZNETSOV, V. V. (2005) [Plastid RNA polymerases]. *Mol Biol (Mosk),* 39**,** 762-75.

MALPICA, R., FRANCO, B., RODRIGUEZ, C., KWON, O. & GEORGELLIS, D. (2004) Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. *Proc Natl Acad Sci U S A,* 101**,** 13318-23.

MANNING, G., PLOWMAN, G. D., HUNTER, T. & SUDARSANAM, S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci,* 27**,** 514-20.

MANUELL, A., BELIGNI, M. V., YAMAGUCHI, K. & MAYFIELD, S. P. (2004) Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome. *Biochem Soc Trans,* 32**,** 601-5.

MARINA, A., WALDBURGER, C. D. & HENDRICKSON, W. A. (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J,* 24**,** 4247-59.

MARTIN, W., RUJAN, T., RICHLY, E., HANSEN, A., CORNELSEN, S., LINS, T., LEISTER, D., STOEBE, B., HASEGAWA, M. & PENNY, D. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.,* 99**,** 12246-12251.

MARTINEZ-HACKERT, E. & STOCK, A. M. (1997) Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol,* 269**,** 301-12.

MASCHER, T., HELMANN, J. D. & UNDEN, G. (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev,* 70**,** 910-38.

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.

MAYFIELD, S. P. & SCHULTZ, J. (2004) Development of a luciferase reporter gene, luxCt, for Chlamydomonas reinhardtii chloroplast. *Plant J,* 37**,** 449-58.

MCKINNEY, E. C., ALI, N., TRAUT, A., FELDMANN, K. A., BELOSTOTSKY, D. A., MCDOWELL, J. M. & MEAGHER, R. B. (1995) Sequence-based identification of T-DNA insertion mutations in Arabidopsis: actin mutants act2-1 and act4-1. *Plant J,* 8**,** 613-22.

MERCHANT, S. S., PROCHNIK, S. E., VALLON, O., HARRIS, E. H., KARPOWICZ, S. J., WITMAN, G. B., TERRY, A., SALAMOV, A., FRITZ-LAYLIN, L. K., MARECHAL-DROUARD, L., MARSHALL, W. F., QU, L. H., NELSON, D. R., SANDERFOOT, A. A., SPALDING, M. H., KAPITONOV, V. V., REN, Q., FERRIS, P., LINDQUIST, E., SHAPIRO, H., LUCAS, S. M., GRIMWOOD, J., SCHMUTZ, J., CARDOL, P., CERUTTI, H., CHANFREAU, G., CHEN, C. L., COGNAT, V., CROFT, M. T., DENT, R., DUTCHER, S., FERNANDEZ, E., FUKUZAWA, H., GONZALEZ-BALLESTER, D.,

GONZALEZ-HALPHEN, D., HALLMANN, A., HANIKENNE, M., HIPPLER, M., INWOOD, W., JABBARI, K., KALANON, M., KURAS, R., LEFEBVRE, P. A., LEMAIRE, S. D., LOBANOV, A. V., LOHR, M., MANUELL, A., MEIER, I., METS, L., MITTAG, M., MITTELMEIER, T., MORONEY, J. V., MOSELEY, J., NAPOLI, C., NEDELCU, A. M., NIYOGI, K., NOVOSELOV, S. V., PAULSEN, I. T., PAZOUR, G., PURTON, S., RAL, J. P., RIANO-PACHON, D. M., RIEKHOF, W., RYMARQUIS, L., SCHRODA, M., STERN, D., UMEN, J., WILLOWS, R., WILSON, N., ZIMMER, S. L., ALLMER, J., BALK, J., BISOVA, K., CHEN, C. J., ELIAS, M., GENDLER, K., HAUSER, C., LAMB, M. R., LEDFORD, H., LONG, J. C., MINAGAWA, J., PAGE, M. D., PAN, J., POOTAKHAM, W., ROJE, S., ROSE, A., STAHLBERG, E., TERAUCHI, A. M., YANG, P., BALL, S., BOWLER, C., DIECKMANN, C. L., GLADYSHEV, V. N., GREEN, P., JORGENSEN, R., MAYFIELD, S., MUELLER-ROEBER, B., RAJAMANI, S., SAYRE, R. T., BROKSTEIN, P., et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science,* 318**,** 245-50.

MIKAMI K, D. S., SULPICE R, TAKAHASHI S, FERJANI A, SUZUKI I AND MURATA N (2003) Histidine Kinases, Hik2, Hik16 and Hik33, in Synechocystis sp. PCC 6803 are Involved in the Torelance of Photosystem II to Environmental Stress. *Plant Cell Physiol,* 44**,** S82.

MIKAMI, K., KANESAKI, Y., SUZUKI, I. & MURATA, N. (2002) The histidine kinase Hik33 perceives osmotic stress and cold stress in Synechocystis sp PCC 6803. *Mol Microbiol,* 46**,** 905-15.

MIKAMI, K., SUZUKI, I. & MURATA, N. (2004) *Plant Responses to Abiotic Stress*, Springer.

MINODA, A., NAGASAWA, K., HANAOKA, M., HORIUCHI, M., TAKAHASHI, H. & TANAKA, K. (2005) Microarray profiling of plastid gene expression in a unicellular red alga, Cyanidioschyzon merolae. *Plant Mol Biol,* 59**,** 375-85.

MINODA, A. & TANAKA, K. (2005) Roles of the transcription factors encoded in the plastid genome of Cyanidioschyzon merolae. IN EST, A. V. D. & BRUCE, D. (Eds.) *Photosynthesis: Fundamental Aspects to Global Perspectives*. Lawrence, Kansas, Alliance Communications Group.

MIZUNO, T. (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of Escherichia coli. *DNA Res,* 4**,** 161-8.

MIZUNO, T. (1998) His-Asp phosphotransfer signal transduction. *J Biochem,* 123**,** 555-63.

MIZUNO, T., WURTZEL, E. T. & INOUYE, M. (1982) Osmoregulation of gene expression. II. DNA sequence of the envZ gene of the ompB operon of Escherichia coli and characterization of its gene product. *J. Biol. Chem.,* 257**,** 13692-8.

MONTGOMERY, B. L. & LAGARIAS, J. C. (2002) Phytochrome ancestry: sensors of bilins and light. *Trends Plant Sci,* 7**,** 357-66.

MOONEY, B. P., DAVID, N. R., THELEN, J. J., MIERNYK, J. A. & RANDALL, D. D. (2000) Histidine modifying agents abolish pyruvate dehydrogenase kinase activity. *Biochem Biophys Res Commun,* 267**,** 500-3.

MORRISON, S. S., MULLINEAUX, C. W. & ASHBY, M. K. (2005) The influence of acetyl phosphate on DspA signalling in the Cyanobacterium Synechocystis sp. PCC6803. *BMC Microbiol,* 5**,** 47.

MOULD, R. M. & ROBINSON, C. (1991) A Proton Gradient Is Required for the Transport of 2 Lumenal Oxygen-Evolving Proteins across the Thylakoid Membrane. *Journal of Biological Chemistry,* 266**,** 12189-12193.

MOUSSATCHE, P. & KLEE, H. J. (2004) Autophosphorylation activity of the Arabidopsis ethylene receptor multigene family. *J Biol Chem,* 279**,** 48734-41.

MULLET, J. E. & KLEIN, R. R. (1987) Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. *EMBO J,* 6**,** 1571-1579.

MURAKAMI, A., FUJITA, Y., A.NEMSON, J. & MELIS, A. (1997a) Chromatic Regulation in Chlamydomonas reinhardtii: Time Course of Photosystem Stoichiometry Adjustment Following a Shift in Growth Light Quality. *Plant Cell Physiol,* 38**,** 188-193.

MURAKAMI, A., KIM, S. J. & FUJITA, Y. (1997b) Changes in photosystem stoichiometry in response to environmental conditions for cell growth observed with the cyanophyte Synechocystis PCC 6714. *Plant and Cell Physiology,* 38**,** 392-397.

NAKAHIRA, Y., BABA, K., YONEDA, A., SHIINA, T. & TOYOSHIMA, Y. (1998) Circadian-regulated transcription of the psbD light-responsive promoter in wheat chloroplasts. *Plant Physiol,* 118**,** 1079-88.

NELSON, N. & BEN-SHEM, A. (2004) The complex architecture of oxygenic photosynthesis. *Nat Rev Mol Cell Biol,* 5**,** 971-82.

NELSON, N. & YOCUM, C. F. (2006) Structure and function of photosystems I and II. *Annu Rev Plant Biol,* 57**,** 521-65.

NEWELL, C. A., BIRCH-MACHIN, I., HIBBERD, J. M. & GRAY, J. C. (2003) Expression of green fluorescent protein from bacterial and plastid promoters in tobacco chloroplasts. *Transgenic Res,* 12**,** 631-4.

NITSCHKE, W. & RUTHERFORD, A. W. (1991) Photosynthetic reaction centres: variations on a common structural theme? *Trends Biochem Sci,* 16**,** 241-5.

NIXON, B. T., RONSON, C. W. & AUSUBEL, F. M. (1986) Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes ntrB and ntrC. *Proc Natl Acad Sci U S A,* 83**,** 7850-4.

OELMÜLLER, R., PESKAN, T., WESTERMANN, M., SHERAMETI, I., CHANDOK, M., SOPORY, S. K., WÖSTEMEYER, A., KUSNETSOV, V., BEZHANI, S. & PFANNSCHMIDT, T. (2001) Novel aspects in photosynthesis gene regulation. IN S.K., S. & R., O. (Eds.) *Signal Transduction in plants*. Dordrecht, Kluwer academic publisher.

OH, J. I. & KAPLAN, S. (2000) Redox signaling: globalization of gene expression. *EMBO Journal,* 19**,** 4237-47.

OHYAMA, K., FUKUZAWA, H., KOHCHI, T., SHIRAI, H., SANO, T., SANO, S., UMESONO, K., SHIKI, Y., TAKEUCHI, M., CHANG, Z., AOTA, S., INOKUCHI, H. & OZEKI, H. (1986) Chloroplast Gene Organization Deduced from Complete Sequence of Liverwort Marchantia-Polymorpha Chloroplast DNA. *Nature,* 322**,** 572-574.

OKADA, M. & HORIE, H. (1979) Diurnal rhythm in the Hill reaction in cell-free extracts of the green alga Bryopsis maxima. *Plant and Cell Physiology,* 20**,** 1403-1406.

PAGET, M. S. & BUTTNER, M. J. (2003) Thiol-based regulatory switches. *Annu Rev Genet,* 37**,** 91-121.

PAITHOONRANGSARID, K., SHOUMSKAYA, M. A., KANESAKI, Y., SATOH, S., TABATA, S., LOS, D. A., ZINCHENKO, V. V., HAYASHI, H., TANTICHAROEN, M., SUZUKI, I. & MURATA, N. (2004) Five histidine kinases perceive osmotic stress and regulate distinct sets of genes in Synechocystis. *J Biol Chem,* 279**,** 53078-86.

PAO, G. M. & SAIER, M. H., JR. (1997) Nonplastid eukaryotic response regulators have a monophyletic origin and evolved from their bacterial precursors in parallel with their cognate sensor kinases. *J Mol Evol,* 44**,** 605-13.

PAWSON, T. & SCOTT, J. D. (2005) Protein phosphorylation in signaling--50 years and counting. *Trends Biochem Sci,* 30**,** 286-90.

PELTIER, G. & COURNAC, L. (2002) Chlororespiration. *Annu Rev Plant Biol,* 53**,** 523-50.

PFANNSCHMIDT, T. & LIERE, K. (2005) Redox regulation and modification of proteins controlling chloroplast gene expression. *Antioxid Redox Signal,* 7**,** 607-18.

PFANNSCHMIDT, T., NILSSON, A. & ALLEN, J. F. (1999a) Photosynthetic control of chloroplast gene expression. *Nature,* 397**,** 625-628.

PFANNSCHMIDT, T., NILSSON, A., TULLBERG, A., LINK, G. & ALLEN, J. F. (1999b) Direct transcriptional control of the chloroplast genes psbA and psaAB adjusts photosynthesis to light energy distribution in plants. *IUBMB Life,* 48**,** 271-276.

PFANNSCHMIDT, T., SCHUTZE, K., BROST, M. & OELMULLER, R. (2001) A novel mechanism of nuclear photosynthesis gene regulation by redox signals from the chloroplast during photosystem stoichiometry adjustment. *Journal of Biological Chemistry,* 276**,** 36125-36130.

PUTHIYAVEETIL, S. & ALLEN, J. F. (2008) Transients in chloroplast gene transcription. *Biochem Biophys Res Commun,* 368**,** 871-4.

QUAIL, P. H. (1991) Phytochrome: a light-activated molecular switch that regulates plant gene expression. *Annu Rev Genet,* 25**,** 389-409.

QUEVILLON, E., SILVENTOINEN, V., PILLAI, S., HARTE, N., MULDER, N., APWEILER, R. & LOPEZ, R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Res,* 33**,** W116-20.

RACE, H. L., HERRMANN, R. G. & MARTIN, W. (1999) Why have organelles retained genomes? *Trends in Genetics,* 15**,** 364-370.

RAVEN, J. A. & ALLEN, J. F. (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol,* 4**,** 209.

REITH, M. & MUNHOLLAND, J. (1993) A High-Resolution Gene Map of the Chloroplast Genome of the Red Alga Porphyra purpurea. *Plant Cell,* 5**,** 465-475.

RENSING, S. A., LANG, D., ZIMMER, A. D., TERRY, A., SALAMOV, A., SHAPIRO, H., NISHIYAMA, T., PERROUD, P. F., LINDQUIST, E. A., KAMISUGI, Y., TANAHASHI, T., SAKAKIBARA, K., FUJITA, T., OISHI, K., SHIN, I. T., KUROKI, Y., TOYODA, A., SUZUKI, Y., HASHIMOTO, S., YAMAGUCHI, K., SUGANO, S., KOHARA, Y., FUJIYAMA, A., ANTEROLA, A., AOKI, S., ASHTON, N., BARBAZUK, W. B., BARKER, E., BENNETZEN, J. L., BLANKENSHIP, R., CHO, S. H., DUTCHER, S. K., ESTELLE, M., FAWCETT, J. A., GUNDLACH, H., HANADA, K., HEYL, A., HICKS, K. A., HUGHES, J., LOHR, M., MAYER, K., MELKOZERNOV, A., MURATA, T., NELSON, D. R., PILS, B., PRIGGE, M., REISS, B., RENNER, T., ROMBAUTS, S., RUSHTON, P. J., SANDERFOOT, A., SCHWEEN, G., SHIU, S. H., STUEBER, K., THEODOULOU, F. L., TU, H., VAN DE PEER, Y., VERRIER, P. J., WATERS, E., WOOD, A., YANG, L., COVE, D., CUMING, A. C., HASEBE, M., LUCAS, S., MISHLER, B. D., RESKI, R., GRIGORIEV, I. V., QUATRANO, R. S.

& BOORE, J. L. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science,* 319**,** 64-9.

RICHLY, E. & LEISTER, D. (2004) An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. *Gene,* 329**,** 11-6.

ROBINSON, V. L., BUCKLER, D. R. & STOCK, A. M. (2000) A tale of two components: a novel kinase and a regulatory switch. *Nat Struct Biol,* 7**,** 626-33.

RONQUIST, F. & HUELSENBECK, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-4.

RUSSELL, M. J. & HALL, A. J. (2006) The onset and early evolution of life. IN KESLER, S. E. & OHMOTO, H. (Eds.) *Evolution of Early Earth's Atmosphere, Hydrosphere, and Biosphere Constraints from Ore Deposits*. Geological Society of America Memoir.

SAITO, H. (2001) Histidine phosphorylation and two-component signaling in eukaryotic cells. *Chem Rev,* 101**,** 2497-509.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989) *Molecular cloning: a laboratory manual,* New York, Cold Spring Harbor Laboratory Press.

SANTOS, J. L. & SHIOZAKI, K. (2001) Fungal histidine kinases. *Sci STKE,* 2001**,** RE1.

SARDIWAL, S., KENDALL, S. L., MOVAHEDZADEH, F., RISON, S. C., STOKER, N. G. & DJORDJEVIC, S. (2005) A GAF domain in the hypoxia/NO-inducible Mycobacterium tuberculosis DosS protein binds haem. *J Mol Biol,* 353**,** 929-36.

SATO, S., SHIMODA, Y., MURAKI, A., KOHARA, M., NAKAMURA, Y. & TABATA, S. (2007) A large-scale protein protein interaction analysis in Synechocystis sp. PCC6803. *DNA Res,* 14**,** 207-16.

SCHEIN, A. I., KISSINGER, J. C. & UNGAR, L. H. (2001) Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Res,* 29**,** E82.

SCHINKMANN, K. & BLENIS, J. (1997) Cloning and characterization of a human STE20-like protein kinase with unusual cofactor requirements. *J Biol Chem,* 272**,** 28695-703.

SCHULTZ, J., MILPETZ, F., BORK, P. & PONTING, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.,* 95**,** 5857-64.

SCHUTZE, K., STEINER, S. & PFANNSCHMIDT, T. (2008) Photosynthetic redox regulation of the plastocyanin promoter in tobacco. *Physiol Plant*.

SELINGER, D. W., SAXENA, R. M., CHEUNG, K. J., CHURCH, G. M. & ROSENOW, C. (2003) Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res,* 13**,** 216-23.

SHIN, D., LEE, E. J., HUANG, H. & GROISMAN, E. A. (2006) A positive feedback loop promotes transcription surge that jump-starts Salmonella virulence circuit. *Science,* 314**,** 1607-9.

SHINOZAKI, K., OHME, M., TANAKA, M., WAKASUGI, T., HAYASHIDA, N., MATSUBAYASHI, T., ZAITA, N., CHUNWONGSE, J., OBOKATA, J., YAMAGUCHI-SHINOZAKI, K., OHTO, C., TORAZAWA, K., MENG, B. Y., SUGITA, M., DENO, H., KAMOGASHIRA, T., YAMADA, K., KUSUDA, J., TAKAIWA, F., KATO, A., TOHDOH, N., SHIMADA, H. & SUGIURA, M. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J,* 5**,** 2043-2049.

SHOUMSKAYA, M. A., PAITHOONRANGSARID, K., KANESAKI, Y., LOS, D. A., ZINCHENKO, V. V., TANTICHAROEN, M., SUZUKI, I. & MURATA, N. (2005) Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in synechocystis. *J Biol Chem,* 280**,** 21531-8.

SIKDAR, S. R., SERINO, G., CHAUDHURI, S. & MALIGA, P. (1998) Plastid transformation in Arabidopsis thaliana. *Plant Cell Reports,* 18**,** 20-24.

SIMPSON, C. L. & STERN, D. B. (2002) The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiol.,* 129**,** 957-66.

SKERKER, J. M., PERCHUK, B. S., SIRYAPORN, A., LUBIN, E. A., ASHENBERG, O., GOULIAN, M. & LAUB1, M. T. (2008) Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell,* 133**,** 1043-1054.

SMALL, I., PEETERS, N., LEGEAI, F. & LURIN, C. (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics,* 4**,** 1581-90.

STEGEMANN, S., HARTMANN, S., RUF, S. & BOCK, R. (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci U S A,* 100**,** 8828-33.

STEPHENSON, K. & HOCH, J. A. (2002) Two-component and phosphorelay signal-transduction systems as therapeutic targets. *Curr Opin Pharmacol,* 2**,** 507-12.

STOCK, A., KOSHLAND, D. E. & STOCK, J. (1985) Homologies between the Salmonella-Typhimurium Chey Protein and Proteins Involved in the Regulation of Chemotaxis, Membrane-Protein Synthesis, and Sporulation. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 7989-7993.

STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. (2000) Two-component signal transduction. *Annu. Rev. Biochem.,* 69**,** 183-215.

STROEBEL, D., CHOQUET, Y., POPOT, J. L. & PICOT, D. (2003) An atypical haem in the cytochrome b(6)f complex. *Nature,* 426**,** 413-8.

SUGIURA, M. (1998) The discovery of the complete sequence of tobacco and rice chloroplast genomes. IN KUNG, S.-D. & YANG, S.-F. (Eds.) *Discoveries in Plant Biology*. Singapore, World Scientific.

SWEM, L. R., GONG, X., YU, C. A. & BAUER, C. E. (2006) Identification of a ubiquinone-binding site that affects autophosphorylation of the sensor kinase RegB. *J Biol Chem,* 281**,** 6768-75.

TAKAI, N., NAKAJIMA, M., OYAMA, T., KITO, R., SUGITA, C., SUGITA, M., KONDO, T. & IWASAKI, H. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc Natl Acad Sci U S A,* 103**,** 12109-14.

TANAKA, T., SAHA, S. K., TOMOMORI, C., ISHIMA, R., LIU, D., TONG, K. I., PARK, H., DUTTA, R., QIN, L., SWINDELLS, M. B., YAMAZAKI, T., ONO, A. M., KAINOSHO, M., INOUYE, M. & IKURA, M. (1998) NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. *Nature,* 396**,** 88-92.

TAYLOR, B. L. (2007) Aer on the inside looking out: paradigm for a PAS-HAMP role in sensing oxygen, redox and energy. *Mol Microbiol,* 65**,** 1415-24.

TAYLOR, B. L. & ZHULIN, I. B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev,* 63**,** 479-506.

THELEN, J. J., MIERNYK, J. A. & RANDALL, D. D. (2000) Pyruvate dehydrogenase kinase from Arabidopsis thaliana: a protein histidine kinase that phosphorylates serine residues. *Biochem J,* 349**,** 195-201.

THOMASON, P. & KAY, R. (2000) Eukaryotic signal transduction via histidine-aspartate phosphorelay. *J Cell Sci,* 113 (Pt 18)**,** 3141-50.

TOUFIGHI, K., BRADY, S. M., AUSTIN, R., LY, E. & PROVART, N. J. (2005) The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J,* 43**,** 153-63.

TSUNOYAMA, Y., ISHIZAKI, Y., MORIKAWA, K., KOBORI, M., NAKAHIRA, Y., TAKEBA, G., TOYOSHIMA, Y. & SHIINA, T. (2004) Blue light-induced

transcription of plastid-encoded psbD gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. *Proc Natl Acad Sci U S A,* 101**,** 3304-9.

TULLBERG, A., ALEXCIEV, A., PFANNSCHMIDT, T. & ALLEN, J. F. (2000) Photosynthetic electron flow regulates transcription of the psaB gene in pea (Pisum sativum L.) chloroplasts through the redox state of the plastoquinone pool. *Plant Cell Physiol.,* 41**,** 1045-1054.

TURKINA, M. V., VILLAREJO, A. & VENER, A. V. (2004) The transit peptide of CP29 thylakoid protein in Chlamydomonas reinhardtii is not removed but undergoes acetylation and phosphorylation. *FEBS Lett,* 564**,** 104-8.

VUILLET, L., KOJADINOVIC, M., ZAPPA, S., JAUBERT, M., ADRIANO, J. M., FARDOUX, J., HANNIBAL, L., PIGNOL, D., VERMEGLIO, A. & GIRAUD, E. (2007) Evolution of a bacteriophytochrome from light to redox sensor. *Embo J,* 26**,** 3322-31.

WAGNER, R., DIETZEL, L., BRAUTIGAM, K., FISCHER, W. & PFANNSCHMIDT, T. (2008) The long-term response to fluctuating light quality is an important and distinct light acclimation mechanism that supports survival of Arabidopsis thaliana under low light conditions. *Planta*.

WAGNER, R. & PFANNSCHMIDT, T. (2006) Eukaryotic transcription factors in plastids--Bioinformatic assessment and implications for the evolution of gene expression machineries in plants. *Gene,* 381**,** 62-70.

WANG, L., SUN, Y. P., CHEN, W. L., LI, J. H. & ZHANG, C. C. (2002) Genomic analysis of protein kinases, protein phosphatases and two-component regulatory systems of the cyanobacterium Anabaena sp. strain PCC 7120. *FEMS Microbiol Lett,* 217**,** 155-65.

WANG, N., SHAULSKY, G., ESCALANTE, R. & LOOMIS, W. F. (1996) A two-component histidine kinase gene that functions in Dictyostelium development. *EMBO J,* 15**,** 3890-8.

WEBER, P., FULGOSI, H., PIVEN, I., MULLER, L., KRUPINSKA, K., DUONG, V. H., HERRMANN, R. G. & SOKOLENKO, A. (2006) TCP34, a nuclear-encoded response regulator-like TPR protein of higher plant chloroplasts. *J Mol Biol,* 357**,** 535-49.

WINTER, D., VINEGAR, B., NAHAL, H., AMMAR, R., WILSON, G. V. & PROVART, N. J. (2007) An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS ONE,* 2**,** e718.

WOLANIN, P. M., THOMASON, P. A. & STOCK, J. B. (2002) Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol,* 3**,** REVIEWS3013.

WU, J., OHTA, N., ZHAO, J. L. & NEWTON, A. (1999) A novel bacterial tyrosine kinase essential for cell division and differentiation. *Proc Natl Acad Sci U S A,* 96**,** 13068-73.

YANO, J., KERN, J., SAUER, K., LATIMER, M. J., PUSHKAR, Y., BIESIADKA, J., LOLL, B., SAENGER, W., MESSINGER, J., ZOUNI, A. & YACHANDRA, V. K. (2006) Where water is oxidized to dioxygen: structure of the photosynthetic Mn4Ca cluster. *Science,* 314**,** 821-5.

YEH, K. C. & LAGARIAS, J. C. (1998) Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc Natl Acad Sci U S A,* 95**,** 13976-81.

ZHANG, Z. G., ZHOU, H. L., CHEN, T., GONG, Y., CAO, W. H., WANG, Y. J., ZHANG, J. S. & CHEN, S. Y. (2004) Evidence for serine/threonine and histidine kinase activity in the tobacco ethylene receptor protein NTHK2. *Plant Physiol,* 136**,** 2971-81.

ZORAGHI, R., CORBIN, J. D. & FRANCIS, S. H. (2004) Properties and functions of GAF domains in cyclic nucleotide phosphodiesterases and other proteins. *Mol Pharmacol,* 65**,** 267-78.