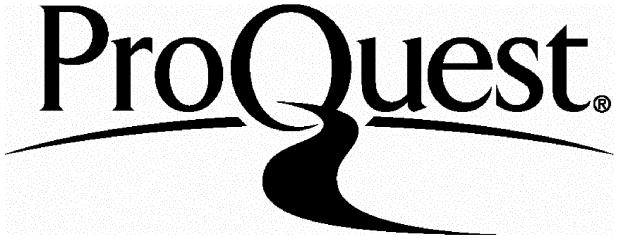
The Transcriptional Apparatus of *Chlamydomonas* Chloroplasts

A thesis submitted for the degree of
Doctor of Philosophy
by

Annette Clare Smith B.Sc. (Hons)

Department of Biology
University College London



Thesis Abstract

The transcriptional apparatus of higher plant chloroplasts is well characterised and consists of a plastid-encoded polymerase (PEP) and a nuclear encoded polymerase (NEP). PEP is dispensable to cell viability. The situation in green algal species, however, is less clear. Chloroplast genes encoding subunits of the PEP have been cloned and sequenced in the green alga *Chlamydomonas reinhardtii* and preliminary reverse-genetic studies suggest that PEP is essential to cell viability, which is in contrast to the situation in higher plants. To investigate this further a series of gene knockouts were constructed using the chloroplast gene *rpo*C2, encoding the β″ subunit of PEP. Results indicate that PEP is essential to *C. reinhardtii* cell viability. In addition, inhibitors of PEP have been used in an *in vivo* transcription assay to try to identify a second RNA polymerase activity in *C. reinhardtii* chloroplasts.

In all higher plant and red algal species so far studied the PEP σ factor is encoded in the nuclear genome. A *C. reinhardtii* nuclear gene (*rpo*D) encoding a putative PEP σ factor has been cloned and partially sequenced. This is the first σ factor cloned from a green algal species. A transcript of ~2.9 kb was detected for the *rpo*D gene by northern analysis.

Finally, epitope tagging technology was developed for chloroplast and bacterial gene products. The *rpo*C2 gene of *C. reinhardtii* was modified to produce a 6x-histidine tagged polypeptide and an attempt was made to purify this polypeptide from *C. reinhardtii* cells using IMAC. In addition, a 3x haemagglutinin (HA) epitope tag was codon optimised for use in *C. reinhardtii* chloroplasts and this epitope was used to tag β-galactosidase in *E. coli*. The protein was detected in a western blot using anti-HA monoclonal antibodies. This epitope will prove useful as a tool to tag *C. reinhardtii* chloroplast proteins.

for Mum and Dad

Acknowledgements

Firstly, I wish to thank my supervisor, Dr Saul Purton for his support and encouragement. Also, I acknowledge the help of Dr Mary Sarcina and Dr Conrad Mullineaux with the fluorescence emission spectroscopy. I owe a great deal to Dr David Stevens for passing on many practical tips, and to Felicity Lown, Nicola Ray and Cleo Bishop for their understanding and friendship. I would also like to acknowledge the help of Dr Sally Hussein, Sam Ali and Wendy Fairclough for proof reading chapters of this thesis.

I would like to acknowledge the Kazusa DNA Research Institute for providing the *rpo*D EST clones, and Prof John Gray for the anti β-galactosidase antibodies.

My deepest thanks go to Adam for his immeasurable support (both emotional and technical) and for understanding. This thesis would never have been completed without it. I would also like to thank Ben and Emma for always listening.

Table of contents

List of figures

List of tables

Abbreviations

µE	microEinstein
µg	microgram
µM	microMolar
µm	micrometre
AGLB	agarose gel loading buffer
Amp	ampicillin
AMV-RT	avian myeloblastosis virus reverse transcriptase
ATP	adenosine 5′-triphosphate
b	bases
β-gal	β-galactosidase
bp	basepairs
BSA	bovine serum albumin
°C	degrees Celsius
cDNA	copy deoxyribonucleic acid
cfu	colony forming units
CIAP	calf intestinal alkaline phosphatase
cm	centimetres
CT	consensus type sequence
CTAB	hexadecyltrimethyl ammonium bromide
CTP	cytidine 5′-triphosphate
cyt b_6f	cytochrome b_6f complex
Da	daltons
dCTP	2′ deoxycytidine 5′-triphosphate
DEPC	diethylpyrocarbonate
ddH2O	double distilled water
dH2O	distilled water
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
ddNTP	dideoxynucleoside 5′-triphosphate

dNTP	deoxynucleoside 5′-triphosphate
EDTA	ethylenediamine tetraacetic acid
EDTA.Na_2	ethylenediamine tetraacetic acid disodium salt
EST	expressed sequence tag
g	gram
GFP	green fluorescent protein
GTP	guanosine 5′-triphosphate
HA	haemagglutinin
HNG	hepes, NaCl and glycerol buffer
HNG-D	hepes, NaCl, glycerol and dodecyl-maltoside buffer
HNG-W	hepes, NaCl, glycerol, dodecyl-maltoside and imidazole buffer
HSM	high salt minimal medium
HTH	helix-turn-helix motif
H_2O	water
lb/in^2	pounds per square inch
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio-β-galactosidase
IR	inverted repeat region
kb	kilobases
kbp	kilobasepairs
kDa	kilodaltons
l	litre
LB	Luria Bertani media
LGLB	Laemmli gel loading buffer
LHCI	light harvesting complex I
LHCII	light harvesting complex II
LSC	large single copy
mg	milligram
min	minute
ml	millilitre
mm	millimetre

mM	milliMolar
MOPS	3-[N-Morpholino] propane-sulfonic acid
mRNA	messenger RNA
mt	mating type
NADP	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine dinucleotide phosphate
NC-II	non consensus type II
NEP	nuclear encoded polymerase
ng	nanogram
Ni^{2+}-NTA	Ni^{2+} - nitrilotriacetic acid
nm	nanometres
NTP	nucleoside 5′-triphosphate
orf	open reading frame
O/N	overnight
P680	primary electron donor in PSII
P700	primary electron donor in PSI
PAGE	polyacrylamide gel electrophoresis
PAR	photosynthetically active radiation
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEP	plastid encoded polymerase
pmol	picomole
P_R	promoter
psi	pounds per square inch
PSI	photosystem I
PSII	photosystem II
pSK	pBluescript (SK-)
PVP	polyvinylpyrrolidone
Q_A	quinone A
Q_B	quinone B
RACE	rapid amplification of cDNA ends
RE	restriction enzyme

Rif	rifampicin
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
RuBPCase	ribulose 1,5-bisphosphate carboxylase
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SLF	sigma like factor
S/N	supernatant
Spc	spectinomycin
SSC	small single copy
SSC	saline sodium citrate
Str	streptomycin
TAC	transcriptionally active chromosome
TAE	tris acetic acid EDTA
TAP	tris acetate phosphate medium
TCA	trichloroacetic acid
Tc^R	tetracycline resistance gene
TEN	tris EDTA NaCl
T_m	melting temperature
Tris	tris(hydroxymethyl)aminoethane
tRNA	transfer RNA
UTP	uridine 5′-triphosphate
UTR	untranslated region
UV	ultraviolet
V	volts
v/v	volume for volume
WT	wild type
w/v	weight for volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase

ycf	hypothetical chloroplast open reading frame

Chapter 1 - Introduction

1.1 Structure, function and origin of the chloroplast

1.1.1 Structure of the chloroplast

Chloroplasts are semi-autonomous organelles found in all plants and eukaryotic algae. Chloroplasts contain a complete set of components needed to carry out photosynthesis. These components are produced by transcription of the organelle genome and translation in the organelle, and by imported proteins transcribed from the nuclear genome and translated in the cytoplasm.

Chloroplasts belong to a supergroup of cell organelles called plastids. Higher plants contain different types of plastids in different plant organs and at various stages of cell development. Plastids can be non-photosynthetic, these include plastid precursors (proplastids and etioplasts), and mature, but specialised plastids such as chromoplasts (which contain carotenoid pigments and are located in fruit, petals and sepals), amyloplasts (which are involved in the synthesis and storage of starch) and elaioplasts (which are oil rich plastids found in oil glands of cacti and epidermal cells of members of the Liliaceae and Orchidaceae) (Emes & Tobin, 1993). The chloroplast is the photosynthetic plastid.

Plant cells vary greatly in both the number and size of their chloroplasts, but in leaves chloroplasts generally occupy about 8% of the total cell volume. Typical higher plant chloroplasts (such as spinach and tobacco) are saucer-shaped, about 4 to 10μm in diameter and 1μm thick. The number of chloroplasts per cell can vary from 1 to 100 depending on the plant and the growth conditions.

Chloroplasts are separated from the rest of the cell cytoplasm by a double outer membrane, the chloroplast envelope. Internally, embedded in the stroma (the chloroplast cytoplasm) the chloroplast contains flattened

thylakoids or lamellae arranged in stacks, called grana. The grana are connected via stroma lamellae, a system of loosely arranged membranes. The chlorophyll-protein complexes that constitute the photosynthetic electron transport chain are embedded in these membranes, the membranes being comprised approximately of half protein and half lipid. The Photosystem II (PSII) complex is found in the grana and the Photosystem I (PSI) complex in the stroma lamellae. Each lamella contains two double layer membranes that are about 5-7nm thick and the space between these membranes is called the lumen. Protons from photosynthetic electron transport are transferred here and used in ATP synthesis (Hall & Rao, 1995).

1.1.2 Functions of the chloroplast

The primary function of the chloroplast is photosynthesis, the reductive biosynthesis of intermediary carbohydrates from CO_2. The chloroplast has other functions such as the biosynthesis of starch, fatty acids, amino acids and pigments and the precursors for secondary metabolism (Neuhaus & Emes, 2000). The chloroplast has its own genetic system and is able to transcribe genes and produce proteins. The photosynthetic and biosynthetic functions will be considered here.

The chemical reactions involved in photosynthesis can be divided into two areas: the photosynthetic electron transfer reactions (light reactions) and the carbon-fixation reactions (dark reactions).

The aim of the light reactions is to generate ATP and NADPH which act as sources of energy and reducing power to convert CO_2 to carbohydrate in the dark reactions. Oxygenic photosynthesis occurs when light is absorbed by chlorophyll and carotenoid pigments and transferred through light harvesting complexes (LHCI or LCHII) via antenna pigments to a chlorophyll *a* dimer. The chlorophyll *a* dimer is located in one of two reaction centres in either PSI

or PSII. The excited chlorophyll *a* dimer releases an electron and starts a series of redox reactions known as the electron transfer chain.

The two photosystems carry out non-cyclic photophosphorylation to produce NADPH and ATP. In PSII the energised chlorophyll *a* dimer, P680, donates electrons to pheophytin. The electron hole in P680 is filled by transfering electrons from H_2O, liberating O_2. Pheophytin transfers electrons to the cytochrome b_6f (cyt b_6f) complex via the plastoquinone pool. PSI, PSII and the cyt b_6f complex are proton pumps, mediating the vectorial transfer of hydrogen ions across the thylakoid membrane and into the lumen. This results in an electrochemical gradient that drives the synthesis of ATP by ATP synthase. The electron is then passed to the energised chlorophyll *a* dimer of PSI (P700) that donates electrons, via the iron-sulphur centre of ferredoxin, to $NADP^+$ to produce NADPH. This process is often referred to as the Z scheme (fig 1.1) (Nugent, 1996).

Carbon fixation occurs by the Calvin-Benson cycle. Three molecules of CO_2 combine with three molecules of ribulose1,5-bisphosphate and, via a complex cyclic series of reactions, ribulose1,5-bisphosphate is regenerated and one molecule of glyceraldehyde 3-phosphate is produced and used to produce sugars, fatty acids and amino-acids. NADPH and ATP are needed in some of the reactions of this cycle (fig 1.2).

Starch biosynthesis occurs in both photosynthetic and non-photosynthetic plastids. Starch synthase and the Branching Enzyme catalyse the transfer of glucosyl units via ADP-glucose to starch. Fatty acid synthesis occurs mainly in the plastids of higher plants. Acetyl CoA carboxylase and fatty acid synthase form long chain fatty acids from acetyl CoA. Amino acids are necessary to the cell to produce protein, transport nitrogen between tissues, synthesise cofactors such as boitin, thiamine, pyrophosphate and coenzyme A and synthesise some pigments such as chlorophyll. Certain amino acids

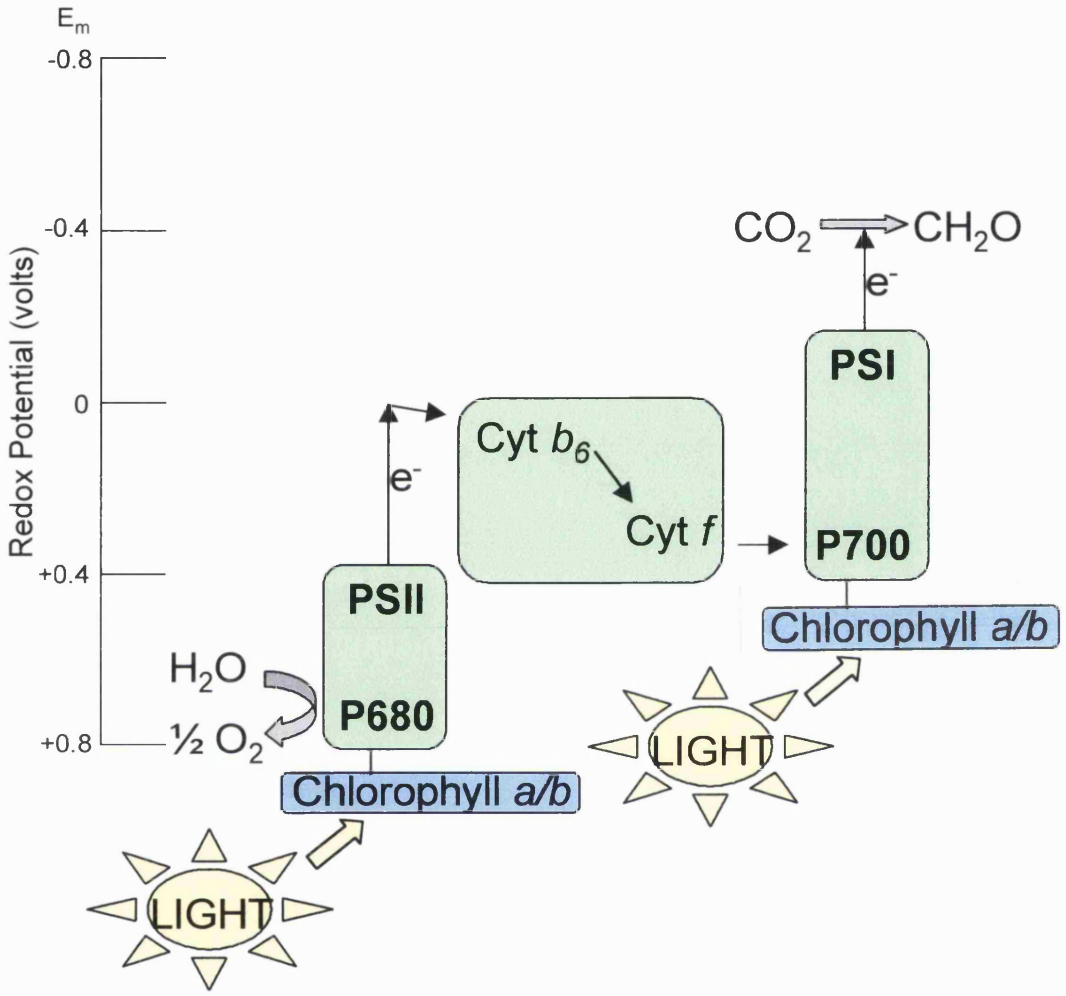


Fig. 1.1 The two light reactions of the "Z scheme" of photosynthesis (from Hall & Rao, 1995)

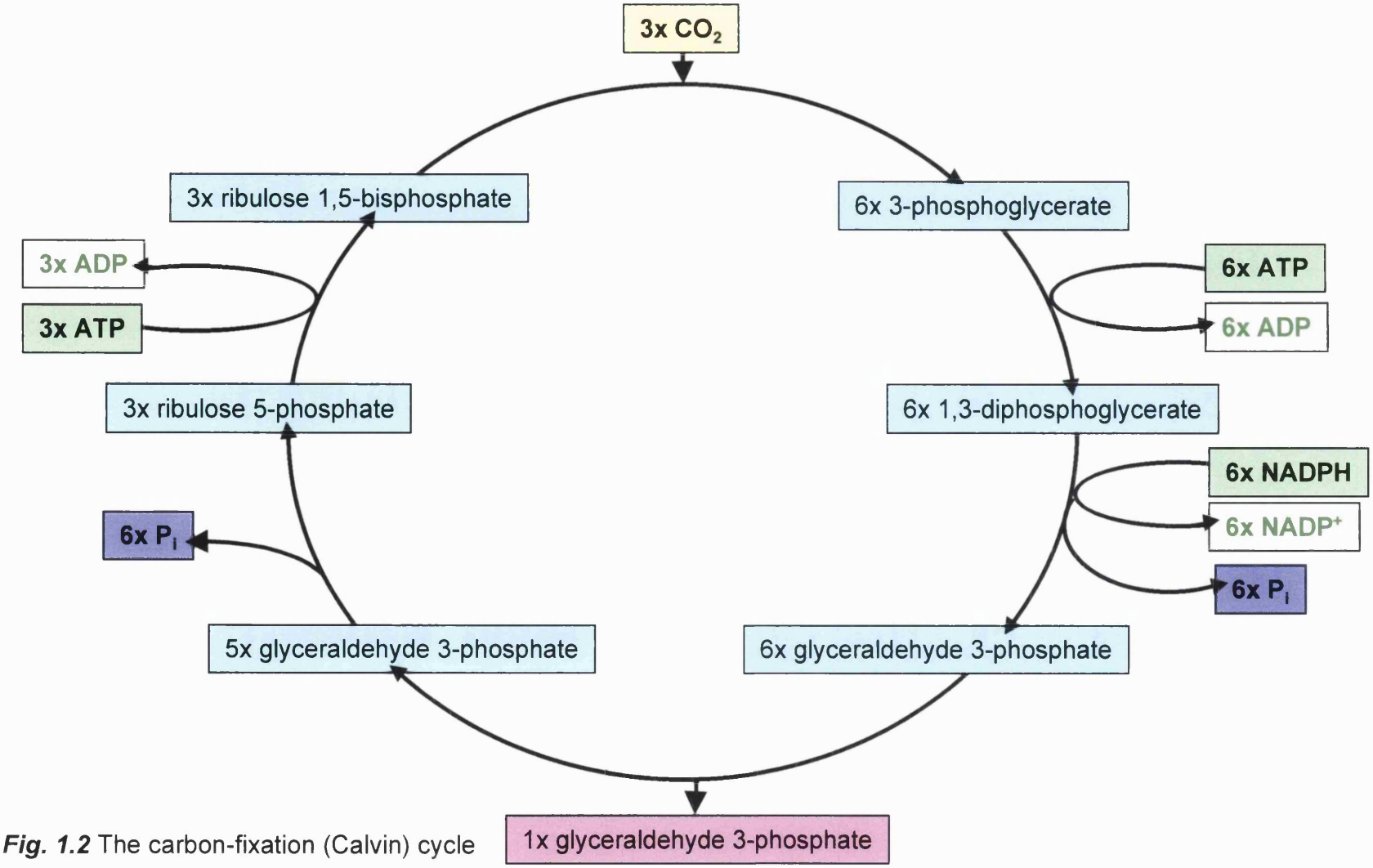


Fig. 1.2 The carbon-fixation (Calvin) cycle

are synthesized in plastids. Plastids also contain an additional pathway of carbohydrate oxidation (Emes & Tobin, 1993).

1.1.3 Evolution of chloroplasts from ancestral cyanobacteria

It is generally agreed that most plastids arose via the invasion of an early eukaryotic cell by a photosynthetic prokaryote (a cyanobacterium) followed by loss of genes to the host cell genome (endosymbiotic gene transfer). This theory was proposed over a century ago, but it was only with the discovery of DNA in the chloroplast followed by the sequencing of complete chloroplast genomes from a range of species that the evidence supporting this theory became overwhelming. Mitochondria are also thought to have arisen by a similar mechanism from purple bacteria (Hess & Borner, 1999).

The sequences of ribosomal RNA (rRNA) genes were initially used to investigate the endosymbiotic theory. These sequences were particularly useful for these studies as they are encoded within both the nuclear and chloroplast genomes (and that of the proposed eubacterial ancestor). It was clear that the chloroplast rRNA sequences are more closely related to cyanobacteria than to those of the "host" cell nuclear genome. Other molecular evidence includes the similarities in arrangement of blocks of genes between the chloroplast and cyanobacteria and the functional similarities in transcription and translation in the organelle and the prokaryote (this will be discussed in detail later in this chapter). Other biochemical and structural data also support this theory (Gray, 1989).

However, whether this endosymbiotic association occurred once (monophyletic origin) or on multiple occasions (polyphyletic origin) is a point of contention. Based on comparisons of biochemical characteristics such as pigment composition it appeared that rhodophyte plastids originated from cyanobacteria as both contain chlorophyll *a* and phycobiliproteins and that chlorophytes originated from prochlorophytes such as *Prochloron didemni* as

both contain chlorophyll *a* and *b*. However subsequent molecular studies do not reinforce these hypotheses. The molecular data, in fact, points to a monophyletic origin. The arrangement of gene clusters and phylogenetic analysis of plastid and cyanobacterial genes is similar enough to point to a single ancestor (Douglas, 1994).

Some algal plastids, however, seem to have arisen from a eukaryote engulfing a eukaryotic algal cell. These plastids are surrounded by more than two membranes (typically three or four); the inner two membranes corresponding to the chloroplast membrane of the endosymbiont, the third membrane corresponds to the plasma membrane of the endosymbiont and the fourth membrane is derived from the food vacuole of the host cell. Following endosymbiosis most of the endosymbionts cellular components (except the plastid) were lost, as were the (principally nuclear) genes. Algae belonging to the cryptomonads and the chloroarachniophyta contain a vestigial nucleus and 80S ribosomes in the periplastidal space, providing strong evidence for this process. The cryptomonads and the chloroarachniophyta, however, have a distinct ancestry as cryptomonad chloroplasts have phycobilins and lack chlorophyll *b* (like red algae) whereas chloroarachniophyta chloroplasts have chlorophyll *b*, but not phycobilins suggesting that it is derived from a green alga (Delwiche, 1999). The apicomplexans have a non-pigmented plastid that is bounded by three or four membranes which is derived from a chloroplast, possibly of a green alga. This plastid is widespread amongst the group, and, although the function is still unknown, it appears to be essential to the cell (Kohler *et al.*, 1997).

1.1.4 Transfer of genes to the nucleus

The nuclear genomes of higher plants and eukaryotic algae contain innumerable genes which are translated in the cell cytoplasm and the products transported into the chloroplast. These genes would have been present in the ancestral cyanobacterial genome which in turn developed into the chloroplast genome.

Over the course of evolutionary time these genes have been translocated from the chloroplast genome to the nucleus. The chloroplast genome encodes only 1-5% of the protein-coding genes of a cyanobacterial genome. The process by which genes are lost from the chloroplast is called endosymbiotic gene transfer, and the pre-requisite for this to occur was the development of apparatus in the chloroplast membrane which allowed the import of proteins produced in the cytoplasm into the organelle. In addition, the transferred genes gain an extra protein sequence (usually at the N-terminus of the protein) which targets the protein to the chloroplast. Machinery must therefore be present to remove this extra protein sequence before the protein will be functional in the organelle.

Although gene regulatory processes in the nucleus are much more complex than those in the chloroplast, this in itself does not explain why genes are transferred to the nucleus to such an extent. The chloroplast is able to tightly regulate transcription and translation, and the nucleus also has control at these levels. For example one of the plastid RNA polymerases (NEP) is nuclear-encoded in higher plants, and the other (PEP) is regulated by nuclear-encoded σ factors. It could be that genes are transferred to the nucleus because plastids reproduce asexually, this means that harmful mutations cannot be recombined out (Müller's Ratchet). Chloroplasts can partially overcome the effects of Müller's ratchet by recombination between genomes within the same chloroplast, but transfer of genes to the nucleus would also prevent some of the harmful effects (Race *et al.*, 1999). Plastid genomes are also AT rich in both coding and non-coding regions. This could be due to the nature of the DNA damage to these genomes. Consequently the high AT content of these genes affects the amino acid composition of plastid proteins. For example plastid genes contain more codons for Phe, Ile, Lys, Asn and Tyr and lack certain codons for Ala, Gly and Pro. These changes in amino acid content may be deleterious for some proteins and a transfer to the nucleus may be advantageous (Howe *et al.*, 2000).

In answering the question "Why are genes transferred to the nucleus?" we raise the question "Why are genes retained by the chloroplast?" It was previously thought that some proteins were too hydrophobic to be transported across the chloroplast membrane, or that codon usage of some chloroplast genes is too different to allow nuclear expression. However the case of the *rbc*L gene, encoding for the large subunit of Rubisco, disproves these theories. *rbc*L is always chloroplast-encoded, despite encoding for a soluble, hydrophilic protein. In addition, the gene can be transferred to the nucleus of tobacco and the transformant will grow phototrophically (reviewed by Race *et al*, 1999). The codons used in the *rbc*L gene can be recognised by cytosolic ribosomes. It is possible that some chloroplast-encoded genes are toxic to the cell in the cytoplasm and ultimately all genes will be lost, as we are not seeing the end point of an evolutionary process. This is certainly the case for hydrogenosomes. This organelle is also of endosymbiotic origin, but it totally lacks a genome (Palmer, 1997). However, in the case of the chloroplast it could be that genes are retained in order to rapidly respond to changes in redox balance. This explains why genes encoding for structural proteins involved in maintaining redox balance across the membranes, along with genes for transcription and translation are retained. Also, most chloroplast genes are lost independently in different lineages, so the remaining genes have been retained in different lineages independently, suggesting a selection pressure at work. It also explains the vast reduction in the plastid genome of parasites such as *E. virginiana* which no longer carry out photosynthesis. Hydrogenosomes have no electron transport across membranes. There is evidence for transcriptional regulation in response to redox state, for example the redox state of plastoquinone controls transcription of genes encoding for components of PSI and PSII (Race *et al.*, 1999).

The exact process by which genes are transferred from the chloroplast to the nucleus is unknown, but it is assumed that initially a copy of the gene is acquired by the nucleus, and that the organelle still retains a functional copy,

consequently the protein is produced by both genomes. Then, the organellar copy becomes degenerate and is ultimately lost. There are no known examples of genes which are functionally produced from both the nuclear and the organellar genome. However, in the case of mitochondria, which are thought to have evolved by the same mechanism, there are examples of newly transferred genes in the nucleus and degenerate genes in the organelle. For example the mitochondrial genome of *Arabidopsis thaliana* contains a defective copy of the *rps*19 gene, and the nucleus contains a functional newly transferred copy (reviewed by Martin & Hermann, 1998). It is assumed that the targeting sequence is acquired almost simultaneously with integration of the gene into the nucleus, allowing rapid degeneration of the organellar gene copy. This is consistent with evidence that acquisition of the targeting sequence is not the rate limiting step in any gene transfer event. However, any period during which the gene is integrated into the nucleus and stably expressed, but not targeted to the organelle, would have a number of consequences for the newly transferred genes as they would be either freed from selective pressures, or subject to new ones. Both of these situations could result in the accumulation of mutations and account, in part, for the differences in organelle-encoded and nuclear-encoded genes.

To suggest that genes are lost from the chloroplast to the nucleus and then the protein products targeted back to the chloroplast would be over-simplifying the process. There are cases which seem to result from loss of genes from the organelle to the nucleus and the protein products never acquire the chloroplast-targeting sequence to allow the protein to function in the organelle. The function is transferred to the cytoplasm and continues to occur at that location. For example, the glycolytic pathway in higher plants appears to be of eubacterial origin, the proteins are of organellar origin and have replaced the endogenous pathway of the original "host". Also, there are chloroplast proteins which appear to result from the transfer of a gene from the mitochondrion to the nucleus, followed by a gene duplication event and the acquisition of a chloroplast targeting sequence onto the protein, which

redirects the protein to the chloroplast. This is then followed by loss of the gene from the chloroplast genome. Usually this occurs for proteins with functions common to both chloroplasts and mitochondria, but has also been suggested to be the origin of the chloroplast nuclear-encoded RNA polymerase (NEP), which is not present in cyanobacteria.

Genes with regulatory functions are transferred to the nucleus more readily than those with enzymatic or structural function. For example the Clp protease consists of two subunits. ClpP is the catalytic subunit and is chloroplast-encoded, and ClpC is the regulatory subunit and is nuclear-encoded (Martin & Herrmann, 1998).

1.2 The chloroplast genome

1.2.1 The chloroplast genetic system

The chloroplast is a semi-autonomous organelle and as such is able to replicate, repair and recombine its DNA, transcribe the genes encoded on its genome and translate the transcripts to produce protein (Rochaix, 1987). This requires proteins that are encoded on the chloroplast genome and produced in the organelle, but also a large number of proteins encoded within the nucleus and targeted to the chloroplast.

DNA replication and repair is essential to the chloroplast genome in order to preserve its DNA. There are two origins of replication on the chloroplast genome of *C. reinhardtii.* Two enzymes with DNA polymerase activity have been isolated from *C. reinhardtii* by separate groups. One is a 180 kDa protein comprised of subunits of 75, 40 and 15 kDa isolated from chloroplasts. The 40 kDa subunit also has exonuclease activity, which is proposed to be involved in proof-reading. The second DNA polymerase consists of two subunits of 116 kDa and 80 kDa and was isolated from cell extracts. Although this protein has biochemical characteristics similar to

higher plant DNA polymerase it is possible that one of these two polymerases is, in fact, the mitochondrial enzyme. DNA gyrases are also involved in DNA replication as DNA gyrase inhibitors inhibit DNA replication and transcription of some genes (Sears, 1998).

The chloroplast DNA of *C. reinhardtii* can be subject to damage by UV irradiation. This damage appears to be repaired by two pathways. A light-dependent photolyase-mediated pathway and a dark repair pathway, although the enzymes involved have not been isolated. Recombination occurs mainly in the Inverted Repeat (IR) region, which is a recombination "hot-spot". It allows copy-correction and "flip-flop" recombination between the two copies of the IR. A RecA like protein is involved in recombination and also in repair (Sears, 1998).

The chloroplast transcription / translation system is strikingly similar to that of the ancestral cyanobacteria and significantly different to that of eukaryotic systems. The chloroplast encoded RNA polymerase is a eubacterial type polymerase and indeed purified *E. coli* RNA polymerase will transcribe cloned chloroplast genes. Chloroplast ribosomes are similar to eubacterial ribosomes in terms of structure and nucleotide sequence. The processes of transcription and translation occur in the chloroplast stroma.

In higher plants there are two RNA polymerases within the organelle which between them are able to transcribe the full set of genes encoded by the plastid genome at all stages of plastid development. One is encoded on the plastid genome and is termed PEP (Plastid Encoded Polymerase) and the other is encoded within the nucleus, NEP (Nuclear Encoded Polymerase). The evidence for the existence of these polymerases and the mechanism of their functions will be considered in detail later in this chapter.

Extensive processing of transcripts occurs in chloroplasts of higher plants and algae. 5′ and 3′ end processing of messenger RNA (mRNA) molecules is

carried out by endo- and exonucleases, which must be imported into the chloroplast. In addition, numerous imported proteins are involved in mRNA stability (Stern & Drager, 1998). Many chloroplast genes contain introns, about half of which have been shown *in vitro* to be removed by self-splicing of the transcript. However, *in vivo* it is thought that nuclear-encoded factors that carry out *trans*-splicing directly or indirectly are present (Herrin & Kuo, 1998). Chloroplast transcripts are sometimes edited (although this has not been demonstrated in any algal species), therefore protein sequences can not always be predicted accurately from DNA sequences (Rochaix, 1992).

The chloroplast translational apparatus consists of 70S ribosomes, transfer RNAs (tRNAs), aminoacyl tRNA synthetases, initiation factors and elongation factors. Genes encoding for rRNA and tRNA are chloroplast encoded and chloroplast ribosomal proteins are chloroplast and nuclear encoded. Other factors involved in chloroplast translation are usually nuclear encoded but can sometimes be found on the chloroplast genome (for example *tuf*A encoding for the translation elongation factor EF-Tu is found on the chloroplast genome of *C. reinhardtii* (Purton, 1995)) providing evidence that these genes have been transferred to the nucleus over the course of evolutionary time.

1.2.2 Structure of the chloroplast genome

Many higher plant chloroplast genomes have been completely sequenced and are strikingly conserved in terms of their size, structure, primary sequence, gene content and organisation. The genome is a double-stranded covalently closed circular molecule of about 120-160 kbp. Plastids contain multiple copies of their genome and the DNA is associated with the inner envelope or the thylakoid membrane as aggregates (nucleoids) of 10-20 copies. The genome is divided by a large inverted repeat region (IR) into a small single copy region (SSC) of 20-24 kbp and a large single copy region (LSC) of 78-100 kbp. The IR is characteristic of chloroplast genomes and can

range in size from 5-76 kbp. All of the genes located in the IR are present as double copies. rRNA genes are always found on the IR, but other genes can be on the IR or on one of the single repeat regions, depending on the size of the IR. The chloroplast genomes of some algae, legumes and conifers lack IRs and in other species such as black pine the IRs are partially lost (Sugiura, 1995). Most genes are organised as multicistronic operons similar to those of the ancestral cyanobacteria. The gene order is highly conserved and any variations can usually be attributed to inversions of large sequences.

Although few green algal chloroplast genomes have been completely sequenced [*Mesostigma viride* (Lemieux *et al.*, 2000), *Nephroselmis olivacea* (Turmel *et al.*, 1999), *Chlorella vulgaris* (Wakasugi *et al.*, 1997) and *Euglena gracilis* (Hallick *et al.*, 1993)] the results indicate that the genomes are less conserved than those of higher plants. They can vary in size from 89-400 kbp and large sequence rearrangements have been identified (Boudreau *et al.*, 1994). Non-green algal plastid genomes differ greatly from those of higher plants and green algae in terms of gene content rather than genome structure (Reardon & Price, 1995).

Non-photosynthetic plastids have genomes that are greatly reduced due to loss of a number of chloroplast genes encoding photosynthetic components. The plastid genomes of the malaria parasite *Plasmodium falciparum* (Wilson *et al.*, 1996) and the parasitic plant *Epifagus virginiana* (Wolfe *et al.*, 1992) retain IRs and the gene order is similar to those of other sequenced plastid genomes, suggesting that deletions rather than rearrangements are responsible for these changes. Fig 1.3 shows a diagram of a typical chloroplast genome.

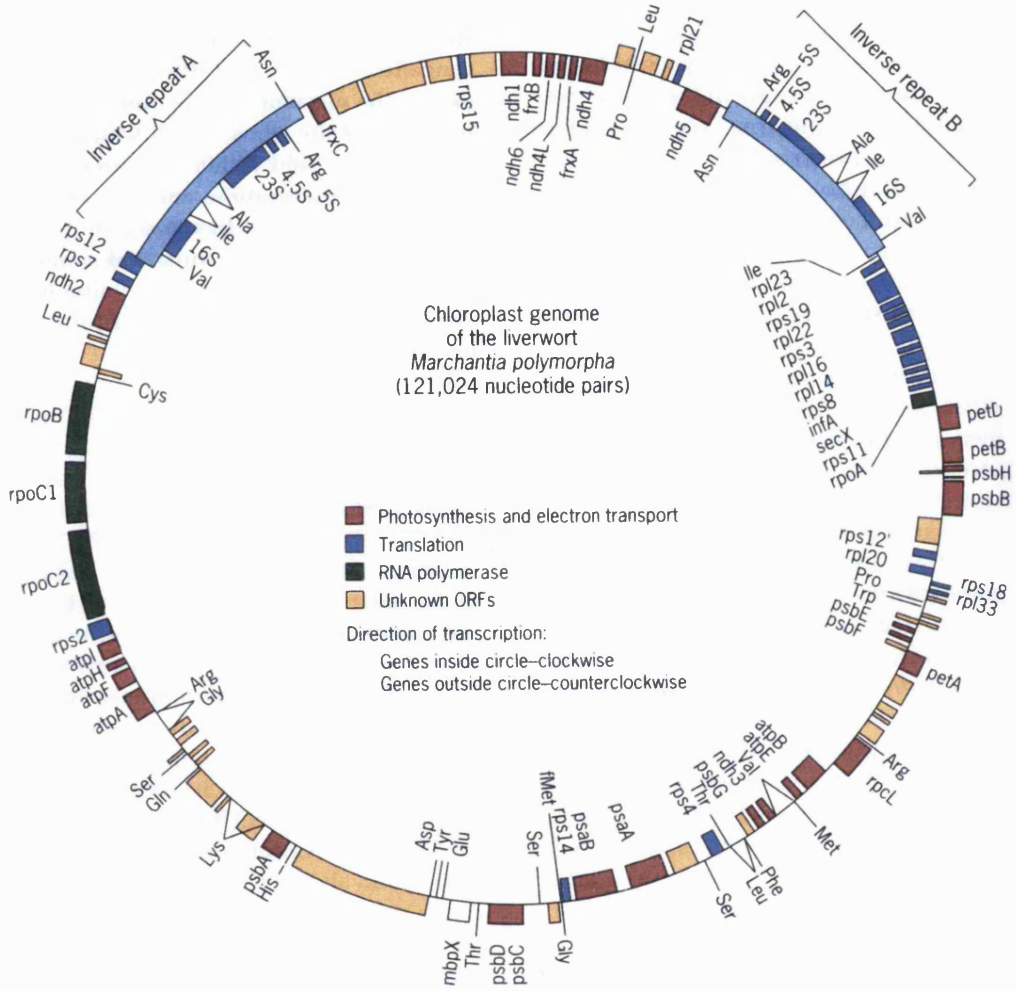


Fig. 1.3 A typical chloroplast genome (taken from Snustad *et al.*, 1997)

Table 1.1 Fully sequenced plastid genomes (adapted from Martin & Herrmann, (1998))

Organism	Genome Size (bp)	Number of protein coding genes	Accession number	Reference
Higher Plants				
Tobacco	155,844	76	S54304	(Shinozaki *et al.*, 1986)
Liverwort	121,024	84	X04465	(Ohyama *et al.*, 1986)
Rice	134,525	76	X15901	(Hiratsuka *et al.*, 1989)
Epifagus virginiana	70,028	21	M81884	(Wolfe *et al.*, 1992)
Pinus thunbergii	119,707	69	D17510	(Wakasugi *et al.*, 1994)
Maize	140,387	76	X86563	(Maier *et al.*, 1995)
Green Algae				
Euglena gracilis	143,170	58	Z11874	(Hallick *et al.*, 1993)
Chlorella vulgaris	150,613	78	AB001684	(Wakasugi *et al.*, 1997)
Nephroselmis olivacea	200,799	127	AF137379	(Turmel *et al.*, 1999)
Mesostigma viride	118,360	135	AF166114	(Lemieux *et al.*, 2000)
Non-Green Algae				
Cyanophora paradoxa	135,599	136	U30821	(Stirewalt *et al.*, 1995)
Porphyra purpurea	191,028	200	U38804	(Reith & Munholland, 1995)
Odontella sinensis	119,704	124	Z67753	(Kowallik *et al.*, 1995)
Guillardia theta	124,524	141		(Douglas & Penny, 1999)
Protists with Plastids				
Plasmodium falciparum	34,682	23	X95275-6	(Wilson *et al.*, 1996)

1.2.3 Chloroplast gene content

The chloroplast genome of higher plants contains 30-36 RNA genes and over 60 protein-coding genes, and, as with chloroplast function, the genes can be divided into two main areas: genes encoding proteins with transcription / translation functions ("housekeeping" functions) and genes encoding proteins with photosynthetic functions. All of the chloroplast rRNA and tRNA genes are encoded by the chloroplast genome. Chloroplast ribosomes consist of about 60 proteins, around 20 of which are chloroplast encoded. Protein complexes with polypeptide subunits encoded within the chloroplast genome include RNA polymerase, PSI and PSII, Rubisco, ATP synthase, cyt b_6f and NADH dehydrogenase (although genes for NADH dehydrogenase are not found on all chloroplast genomes). All of these complexes also consist of many nuclear-encoded polypeptides. In addition, the chloroplast genome contains a number of open reading frames (orfs) of unknown function. Table 1.2 is a summary of all of the chloroplast genes and their gene products that have currently been identified.

Algal chloroplast genomes, being more diverse than higher plant chloroplast genomes, sometimes contain additional genes that are not found on higher plant chloroplast genomes. For example, the chloroplast genome of the red alga *Porphyra purpurea* contains an additional 70 genes compared to higher plants. Some of these additional genes are involved in biosynthetic functions including synthesis of amino acids, fatty acids, pigments and thiamine (Reith & Munholland, 1995). In higher plants these genes must have been transferred to the nucleus.

Table 1.2 Plastid genes and gene products (adapted from Stoebe et al., (1998)) (genes are included if they are present in five or more completely sequenced plastid genomes)

Gene(s)	Function	Gene(s)	Function
***acc*D**	Acetyl-CoA carboxylase carboxytransferase	***pet*A, *pet*B, *pet*D, *pet*G, *pet*L, *pet*N**	Cytochrome b_6/f
***atp*A, *atp*B, *atp*E, *atp*F, *atp*H, *atp*I**	ATP Synthase	***psa*A, *psa*B, *psa*C, *psa*I, *psa*J, *psa*M**	PSI
***ccs*A (*ycf*5)**	Haem attachment to plastid cyt *c*	***psb*A, *psb*, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T**	PSII
***cem*A (*ycf*10)**	Envelope membrane protein	***rbc*L**	Rubisco
***chl*B, *chl*I, *chl*L, *chl*N,**	Protochlorophyllide reductase	***rpl*2, *rpl*5, *rpl*6, *rpl*12, *rpl*14, *rpl*16, *rpl*19, *rpl*20, *rpl*21, *rpl*22, *rpl*23, *rpl*32, *rpl*33, *rpl*36**	Ribosomal protein (large)
***clp*P**	Clp protease	***rpo*A, *rpo*B, *rpo*C1, *rpo*C2**	RNA polymerase
***inf*A**	Translation initiation factor	***rps*2, *rps*3, *rps*4, *rps*5, *rps*7, *rps*8, *rps*9, *rps*11, *rps*12, *rps*14, *rps*15, *rps*16, *rps*17, *rps*18, *rps*19**	Ribosomal protein (small)
***mat*K (*ycf*14)**	Intron maturase	***tuf*A**	Translation elongation factor Tu
***ndh*A, *ndh*B, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K**	NADH-plastoquinone oxidoreductase	***ycf*3, *ycf*4, *ycf*6, *ycf*9, *ycf*12, *ycf*24,**	ORFs of unknown function

1.3 *C. reinhardtii* and chloroplast reverse genetics

1.3.1 C. reinhardtii *as a model organism*

C. reinhardtii has been referred to as the "green yeast" as it is unicellular, cheap and easy to grow on nutrient agar or in liquid culture. Conventional genetic analysis can be easily carried out on all three genomes (nuclear, chloroplast and mitochondrial) as nuclear genes segregate in a Mendelian 2:2 ratio whereas chloroplast and mitochondrial genes are uniparentally inherited from mating-type + (mt+) and mating-type – (mt-) parents, respectively. All three genomes are transformable. As *C. reinhardtii* is photosynthetic and flagellated it is a preferred model system for studies on these processes and their associated structures (Nickelsen & Kuck, 2000). An advantage of using *C. reinhardtii* in the study of photosynthesis is that cells will grow heterotrophically in the presence of acetate, allowing the organism to dispense with photosynthetic function. Photosynthetic deficient mutants can easily be isolated and maintained (Harris, 1989). Cells can be grown phototrophically (in the light with CO_2 as the only carbon source), heterotrophically (in the dark with acetate) or mixotrophically (in the light with acetate). The cell division cycles of *C. reinhardtii* can be synchronized by subjecting the cells to alternate light and dark cycles under photosynthetic growth conditions (Rochaix, 1995).

1.3.2 Structure of the C. reinhardtii *chloroplast*

C. reinhardtii cells contain just one cup-shaped chloroplast that occupies about 40% of the volume of the cell and may partially surround the nucleus. It is a highly organised body, bounded by a double envelope membrane, and containing stacks of lamellae which are the site of photosynthesis. Substructures associated with other chloroplast functions are also present in the stroma such as the eyespot, and the pyrenoid (Olive & Wollman, 1998).

In common with vascular plants, *C. reinhardtii* chloroplasts have a basic chloroplast envelope that consists of two membranes. The inner envelope membrane regulates the transport of metabolites into and out of the chloroplast and has a distinct set of polypeptides to the outer envelope membrane that is permeable to many low molecular weight substances. The envelope plays an essential role in the import of nuclear-encoded chloroplast-targeted proteins and in the regulation of transport of metabolites between the stroma and cytosol. It is also involved in lipid biosynthesis.

The eyespot of *C. reinhardtii* is situated beneath the chloroplast envelope. As *C. reinhardtii* is a flagellated alga, the eyespot supports a photoreceptor, which allows phototaxis. The eyespot may also be involved in reflecting and intensifying light of a specific spectral range or may act as a shading device. The pyrenoid is a differentiated region of the chloroplast stroma, which is thought to be involved in starch biosynthesis, as it is surrounded by starch-containing bodies. However, the main component of purified pyrenoids is active Rubisco. The exact function of the pyrenoid is unknown, but these data suggest a role in the coupling of carbon fixation and starch metabolism.

The thylakoid membranes of *C. reinhardtii* differ greatly from those of vascular plants. The membranes still form long, flat vesicles, known as discs, which are arranged in stacks of two to ten in wild-type chloroplasts to form grana. The grana are spatially close and may be linked via a given disc. In vascular plants, short segments of the grana, consisting of many discs are linked via stroma lamellae, a network of long single discs. All the complexes involved in the photosynthetic electron transport chain are located in the thylakoid membrane and are distributed between stacked and unstacked membranes. The PSII centres and their peripheral antennae (LHCII) are found in the stacked regions of the membranes and the PSI centres, their peripheral antennae (LHCI) and the ATP synthase are located in the unstacked regions. The cyt b_6f complex is found in both domains although all of these complexes are laterally mobile.

Fig 1.4 shows a diagram of the *C. reinhardtii* cell and chloroplast.

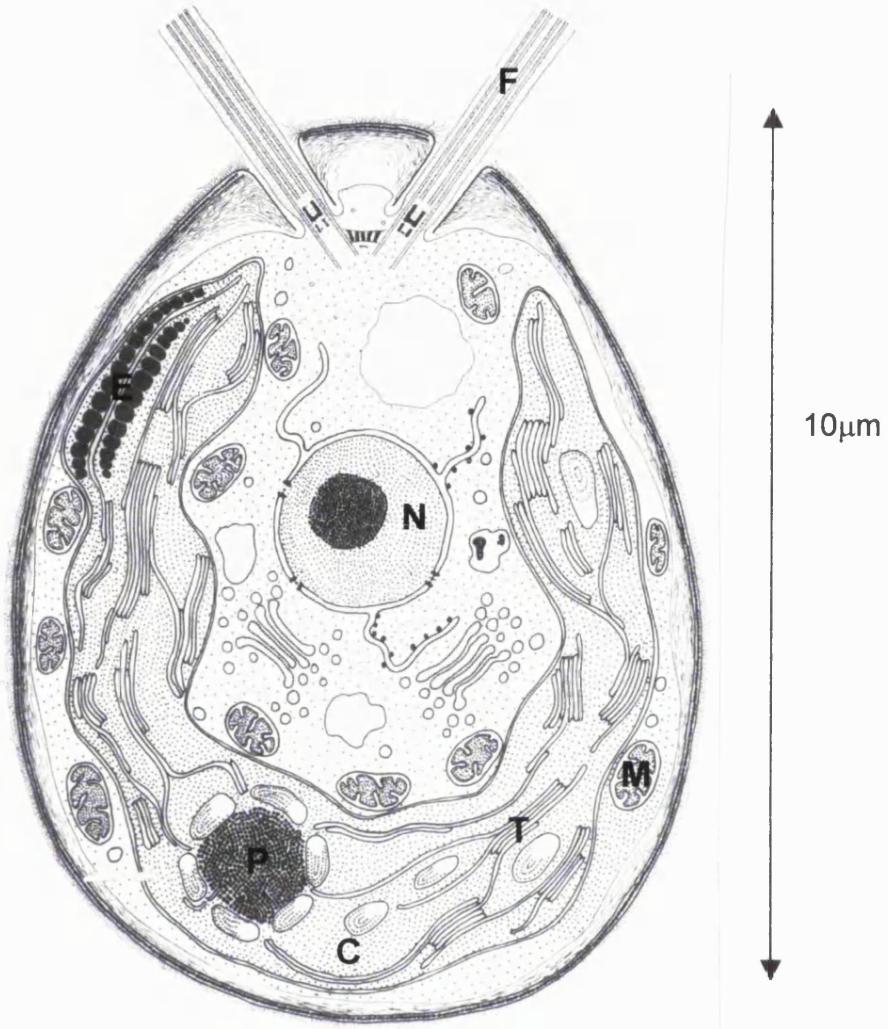
1.3.3 Structure of the C. reinhardtii chloroplast genome

The *C. reinhardtii* chloroplast genome is 196 kbp in size and there are around 80 copies of the genome per chloroplast. The IRs are 21 kbp and the single copy regions are roughly equal in size. The genome is almost completely sequenced. The gene content is essentially the same as that of higher plants, the differences between the *C. reinhardtii* and higher plant chloroplast genomes lie in the organisation of the genes. The genes are arranged into operons less frequently and the gene order differs to that of higher plants (Purton, 1995).

A striking structural feature of the chloroplast genome of *C. reinhardtii* is that the *psa*A gene encoding for the PSI-A protein is split into three exons and located at different sites on the genome. The exons are transcribed independently and the transcripts assembled by *trans*-splicing involving another chloroplast-encoded gene, *tsc*A (Rochaix, 1992).

Fig 1.5 shows the *C. reinhardtii* chloroplast genome.

A number of genes encoded within the chloroplast genome of *C. reinhardtii* are unusual in that they are very large and their deduced amino-acid sequences show blocks of sequences homologous to the equivalent higher plant protein, separated by extra coding sequences which are not spliced at the RNA level and are in-frame with the rest of the gene. Examples of such genes include *rpo*C2 encoding for the β″ subunit of the PEP, *cem*A an envelope membrane protein, *clp*P encoding for the catalytic subunit of the ATP-dependent Clp protease (Huang *et al.*, 1994) and *rps*3 a ribosomal protein (Turmel & Otis, 1994). It has been proposed that these extra coding regions may represent "protein introns" or "inteins". Protein splicing is defined as the excision of an intervening protein sequence (the intein) from a protein



Key	
C	Chloroplast
E	Eyespot
F	Flagellum
M	Mitochondrion
N	Nucleus
P	Pyrenoid
T	Thylakoid membranes

Fig. 1.4 The *C. reinhardtii* cell

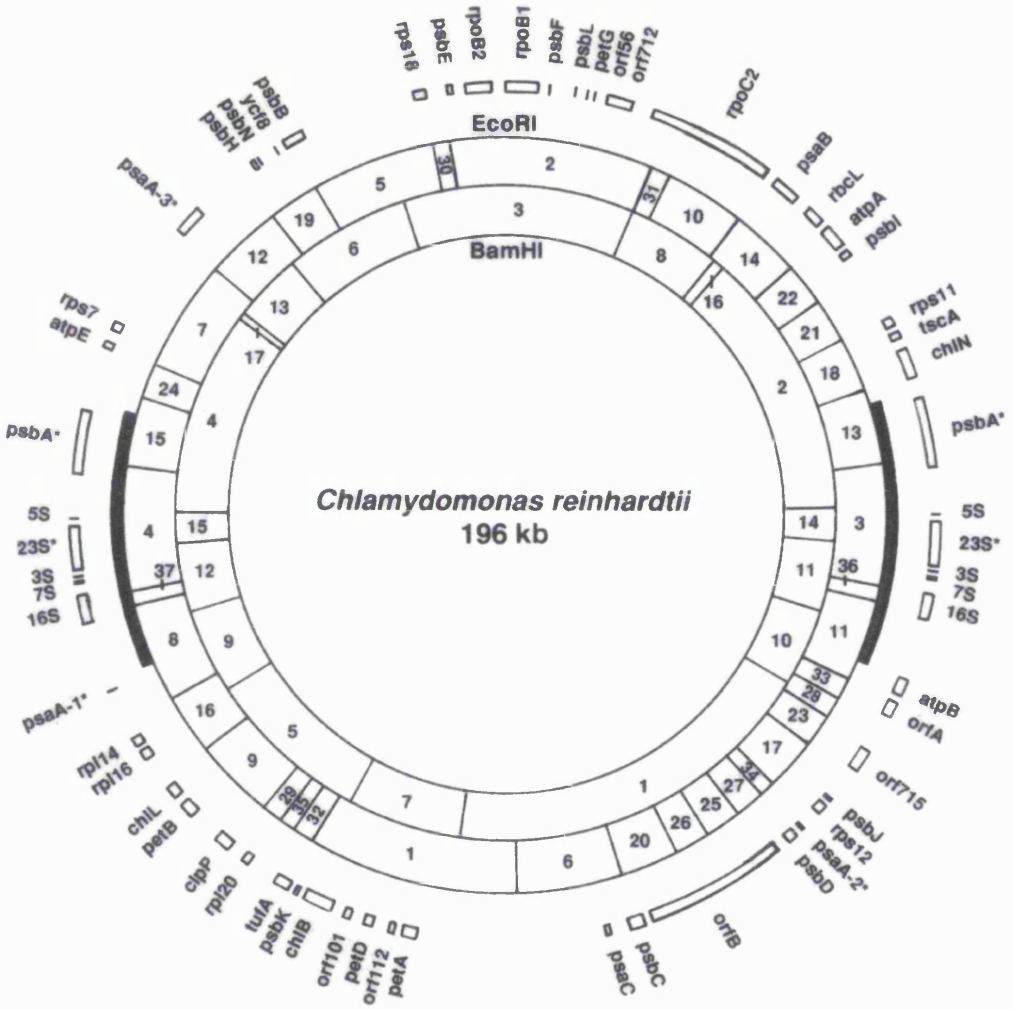


Fig. 1.5 The chloroplast genome of *C. reinhardtii*

precursor along with concomitant ligation of the flanking protein fragments (the exteins) to form a mature protein and a free intein (Perler *et al.*, 1994). A native peptide bond is formed between the ligated exteins (Cooper & Stevens, 1993) and it is the extein ligation which differentiates protein splicing from other forms of autoproteolysis. Protein splicing has been demonstrated in a range of species such as *Saccharomyces cerevisiae* (Kawasaki *et al.*, 1997), *Chlamydomonas eugametos* (Wang & Liu, 1997), *Synechocystis* (Mathys *et al.*, 1999), *Mycobacterium tuberculosis* (Davis *et al.*, 1992), *M. leprae* (Davis *et al.*, 1994), *Thermococcus litoralis* and *Pyrococcus sp.* (Hodges *et al.*, 1992). Inteins from different organisms have a set of conserved amino acids at both splice junctions. Cysteine or serine are always found at the N-extein (the extein at the N-terminal end of the protein) splice junction and histidine, asparagine and cysteine, serine or threonine at the C-extein splice junction (Koonin, 1995). The mechanism of protein splicing initially involves the cleavage of the N-extein splice junction which leads to the formation of a branched intermediate. The branched intermediate has an ester link with the C-terminal carboxylate of the N-extein and the side chain hydroxyl of the serine at the C-extein splice junction. The side chain amino group of the asparagine attacks the asparagine peptide carbonyl and an intramolecular succinimide ring forms. The peptide bond between the intein and the N-extein which is ester linked to the serine residue of the C-extein is cleaved. Finally the N-extein which is ester linked to the serine residue of the C-extein undergoes an O-N shift to form the mature protein, both exteins being linked by normal peptide bonds (Xu *et al.*, 1993).

1.3.4 C. reinhardtii *chloroplast gene content*

As for higher plants, the chloroplast genome of *C. reinhardtii* contains all the tRNA and rRNA genes and about one third of the genes for ribosomal protein subunits. The photosynthetic gene content is similar to that of higher plants but significant differences do exist. The gene encoding for the translation elongation factor EF-Tu (*tuf*A) is present in the *C. reinhardtii* chloroplast

genome but is located in the nuclear genome of higher plants. Three genes involved in the reduction of protochlorophyllide to chlorophyllide (*chl*B, *chl*L and *chl*N) are present in *C. reinhardtii*, as they are in all species that can synthesise chlorophyll in the dark. These genes are absent from the chloroplast genomes of species that are dependent on light (eg angiosperms). Genes which are absent from the *C. reinhardtii* chloroplast genome include the *rpo*A gene encoding for the α subunit of the PEP. This gene is absent from the chloroplast genome of *Euglena gracilis* (Hallick *et al.*, 1993) and the plastid genomes of two non-photosynthetic protists *Plasmodium falciparum* (Wilson *et al.*, 1996) and *Toxoplasma gondii*. In the case of *P. falciparum* the *rpo*A gene has been identified in the nucleus (Sato *et al.*, 2000). The *ndh* genes encoding for NADH dehydrogenase are also absent. These genes are also absent from the *E. gracilis* and *C. vulgaris* (Wakasugi *et al.*, 1997) chloroplast genome, but present in the *N. olivacea* (Turmel *et al.*, 1999) chloroplast genome. These genes reside in the nucleus in *C. reinhardtii* as *rpo*A is an essential subunit of the PEP and there is biochemical evidence for a chlororespiratory complex in the *C. reinhardtii* chloroplast.

Recently the *psa*I gene, encoding for a subunit of PSI has been cloned and sequenced from the nuclear genome of *C. reinhardtii*. This gene is located in the chloroplast genome of all completely sequenced species except the green algae *C. vulgaris* (Wakasugi *et al.*, 1997) and *E. gracilis* (Hallick *et al.*, 1993) (K. Ali unpublished results).

1.3.5 Chloroplast transformation

Chloroplast transformation was first achieved in *C. reinhardtii* cells using particle gun bombardment, also known as the biolistic method (Boynton *et al.*, 1988). Gold or tungsten microparticles are coated with the transforming DNA and are delivered into the chloroplast, across the cell wall and three membranes, by particle bombardment of a lawn of algal cells. The

bombardment is effected by acceleration of the microparticles by gun powder explosion or expansion of compressed gas. The transforming DNA integrates into the chloroplast genome by homologous recombination, allowing the targeted deletion or mutation of chloroplast genes (Nickelsen & Kuck, 2000). In the initial experiments a photosynthetic-deficient mutant resulting from a deletion in the *atp*B gene was "rescued" by transformation with a functional copy of the gene, and selection by phototrophic growth (Boynton *et al.*, 1988). Chloroplasts can be transformed by other methods such as polyethylene glycol (PEG) mediated transformation of protoplasts or microinjection of plastids (Heifetz, 2000) but the biolistic method remains the most commonly used.

Later a variant of the 16S rRNA gene which confers resistance to the antibiotics streptomycin and spectinomycin and a variant of the 23S rRNA gene which confers resistance to erythromycin was used to transform wild-type *C. reinhardtii* cells. The resulting antibiotic resistant transformants were shown to result from integration of the variant genes (Newman *et al.*, 1990). The 16S rRNA gene was later exploited in a co-transformation to disrupt the *atp*B gene. Wild-type *C. reinhardtii* cells were transformed with the spectinomycin resistant variant of the 16S rRNA gene along with a deleted version of the *atp*B gene. The resulting transformants were selected for on the basis of spectinomycin resistance, but were also non-photosynthetic as a result of the *atp*B deletion (Kindle *et al.*, 1991). This method can be used to efficiently disrupt non-essential chloroplast genes.

Chloroplast transformation can be used to carry out reverse-genetics on genes with unknown functions to identify a null phenotype, or to introduce site directed changes in genes with known functions to identify functionally important regions of the gene / protein (Rochaix, 1997).

1.3.6 Chloroplast selectable markers

The first, and most widely used dominant selectable marker for chloroplast transformation in *C. reinhardtii* is the *aad*A cassette. Goldschmidt-Clermont (1991) constructed a series of expression vectors consisting of transcription and translation signals from chloroplast genes. The bacterial *aadA* gene coding for the aminoglycoside 3′ adenyl transferase was inserted into the expression vectors and used to transform wild-type *C. reinhardtii.* Stable transgenic expression of the protein was confirmed by the spectinomycin and streptomycin resistance of the resulting transformants. The *aad*A cassette was used to disrupt photosynthetic genes such as *tsc*A and *psa*C and a gene of unknown function, orf472. The most commonly used form of the *aad*A cassette is driven from the promoter of the *atp*A gene and at the 3′ end has the untranslated region 3′ (UTR) of the *rbc*L gene. This study also represented the first expression of a foreign gene in chloroplasts (Goldschmidt-Clermont, 1991).

Chloroplast reverse-genetics in *C. reinhardtii* has been hampered by the lack of a second dominant selectable marker. Recently a selectable marker based on the bacterial *aphA-6* gene, which encodes an aminoglycoside phosphotransferase, has been developed by Bateman and Purton (2000). The *aphA-6* gene confers resistance to kanamycin and amikacin. The *aphA-6* marker is driven from the promoter of the *psb*A gene and also contains the *rbc*L 3′ UTR. This cassette has been successfully used to disrupt the *psb*H gene by selection on kanamycin. This marker allows a great many new chloroplast mutants to be made. For example, double mutants with modifications to two separate genes on the chloroplast genome can be made. Null mutants with a gene deletion using one marker can be re-transformed with modified versions of the gene using the second marker. This allows modifications to photosynthetic genes which do not allow phototrophic growth and modifications to non-photosynthetic genes. Previously re-transformation

would have relied on selection by phototrophic growth (Bateman & Purton, 2000).

1.3.7 Disruption of non-essential and essential chloroplast genes

The disruption of a chloroplast gene results in a homoplastic or heteroplastic transformant. In a homoplastic transformant all copies of the chloroplast genome contain a disrupted copy of the gene. This can be readily achieved after a few rounds of single colony isolation in the presence of a selective pressure (such as an antibiotic) when disrupting non-essential genes. In a heteroplastic transformant, however, the cell maintains a mixture of both disrupted and wild-type gene copies. This is because the cell requires both wild-type gene copies and the selectable marker for growth on the selective medium (Rochaix, 1995).

1.4 Transcription of chloroplast genes

The enzymes involved in transcription of chloroplast genes in higher plants have been well characterised in recent years. At least two RNA polymerases transcribe chloroplast encoded genes. There is a multi-subunit "eubacterial" type RNA polymerase, the core subunits of which are encoded within the chloroplast genome (PEP - Plastid Encoded Polymerase) and a single-subunit "phage" type polymerase, which is encoded within the nucleus (NEP - Nuclear Encoded Polymerase). This has been determined recently using molecular-genetic studies, but the existence of two polymerases has long been inferred from biochemical studies.

1.4.1 The discovery of an RNA polymerase activity in chloroplasts

A DNA dependent RNA polymerase activity was initially identified in chloroplasts from work on highly purified preparations of broad bean chloroplasts. This activity incorporates ^{14}C-ATP into RNA and is dependent

upon all four NTPs. In common with RNA polymerases from other organisms the activity is sensitive to actinomycin D (a general RNA polymerase inhibitor) (Kirk, 1964).

A DNA dependent RNA polymerase activity similar to that of bacterial systems was identified in tobacco chloroplasts. The RNA polymerase showed properties similar to RNA polymerases found in both animal and bacterial systems, such as an absolute requirement for all four NTPs and Mg^{2+}, and sensitivity to actinomycin D. In one respect, however, the chloroplast enzyme differed significantly from the nuclear enzyme. The activity of the nuclear enzyme was highly repressed due to the presence of histones on the nuclear DNA. Chloroplast DNA lacks histones, so the chloroplast RNA polymerase was able to synthesise long chains of RNA. This is more similar to a bacterial system (Tewari & Wildman, 1969).

Initial studies, carried out before the cloning and sequencing of chloroplast genes, focused on the products of the chloroplast RNA polymerase in order to elucidate the function of the chloroplast genetic system. It was tentatively suggested by Spencer and Whitfield that chloroplast RNA polymerase of spinach can synthesise mRNA. This was based on similarities in the sedimentation patterns between mRNAs and the chloroplast RNA polymerase RNA product. In this study there was no correlation seen between the chloroplast RNA polymerase RNA product and tRNAs or rRNAs (although this could be due to a slower rate of synthesis of tRNAs and rRNAs) (Spencer & Whitfield, 1967). Although it had been well established that the chloroplast contained ribosomes it was unclear whether chloroplast proteins were encoded within the chloroplast genome or whether nuclear encoded mRNAs were imported into the chloroplast and translated in the organelle. This study indirectly suggested that the chloroplast genome encoded chloroplast proteins.

1.4.2 Properties of the TAC and the soluble RNA polymerase

Early studies involving the biochemical isolation of RNA polymerase activities from a number of plants and algae including maize (Bottomley *et al.*, 1971), wheat (Polya & Jagendorf, 1971), *Euglena gracilis* (Hallick *et al.*, 1976), spinach (Briat *et al.*, 1979) and pea (Tewari & Goel, 1983) suggested that there were two RNA polymerase activities present in chloroplasts. It was concluded that the RNA polymerase activities were distinct on the basis of their different solubilities. One enzyme could be solubilised from lysed chloroplasts by the addition of 0.5M salt and was therefore termed the soluble enzyme. The other enzyme was tightly bound to the chloroplast genome and the thylakoid membrane, forming the transcriptionally active chromosome (TAC) which could be solubilised by the addition of 1% Triton X-100 (Bottomley, 1970).

A soluble RNA polymerase with a molecular weight of around 500 kDa from chloroplastswas purified from maize seedlings. This RNA polymerase was insensitive to both α-amanitin, an inhibitor of eukaryotic nuclear RNA polymerases and rifamycin, an inhibitor of eubacterial RNA polymerase (Bottomley *et al.*, 1971).

A soluble chloroplast RNA polymerase isolated from wheat leaves showed a similar susceptibility to these inhibitors. The enzyme was inhibited by actinomycin D, but was insensitive to rifampicin, streptolydigin (an inhibitor of eubacterial RNA polymerase) and α-amanitin (Polya & Jagendorf, 1971). Early work on the chloroplast RNA polymerase of spinach isolated a transcriptionally active DNA-protein complex which is also rifampicin resistant (Briat *et al.*, 1979). Subsequent work showed that this DNA-protein complex consisted of eight polypeptide subunits ranging in size from 15-80 kDa (Briat & Mache, 1980).

Later work on a transcriptionally active DNA-protein complex isolated from pea resolved the subunit composition further. It was determined that seven polypeptides of 180, 140, 110, 95, 65, 47 and 27 kDa were present. Again this enzyme was insensitive to rifampicin and α-aminitin (Tewari & Goel, 1983).

A DNA-protein complex termed the transcriptionally active chromosome (TAC) was also isolated from chloroplasts of the green alga *E. gracilis*. Again, the RNA polymerase activity associated with this complex is insensitive to rifampicin, streptolydigin and α-amanitin, but sensitive to actinomycin D (Hallick *et al.*, 1976). This enzyme has been well characterised. Rushlow and co-workers showed that the TAC can selectively transcribe from the endogenous chloroplast DNA template and that TAC can initiate and elongate RNA molecules *in vitro*. Sequences of RNA molecules synthesised by TAC *in vitro* are the same as RNA molecules synthesised *in vivo*. The TAC of *E. gracilis* transcribes predominately rRNA genes (Rushlow *et al.*, 1980) even when the TAC is purified further to consist of just three polypeptides (Narita *et al.*, 1985). A soluble RNA polymerase which is insensitive to α-amanitin and rifampicin was also isolated from *E. gracilis* chloroplasts and this can transcribe cloned chloroplast tRNA genes (Gruissem *et al.*, 1983). The two RNA polymerase activities differ in that the soluble RNA polymerase is sensitive to salt and heparin whereas the TAC is insensitive to high concentrations of both. This division of labour between the two polymerase activities with the TAC transcribing rRNA genes and the soluble enzyme transcribing tRNA genes was also seen in spinach (Gruissem *et al.*, 1983).

Despite these findings later work by Krupinska and Falk found that the TAC isolated from barley chloroplasts was able to transcribe all classes of plastid genes *in vitro* (Krupinska & Falk, 1994).

Work by Apel and Bogorad focused on the effect of light on the maize chloroplast soluble RNA polymerase. It was observed that following 16 hours of illumination of dark grown seedlings, the soluble RNA polymerase activity increases 3-4 fold relative to dark grown seedlings which have been illuminated for 2 hours. RNA polymerase activity was measured as rate of synthesis of RNA molecules and this increase was not due to an increase in the amount of RNA polymerase, nor to changes in the purified enzyme. It was suggested that other factors which interact with the enzyme or the DNA template might be responsible for this increase in activity (Apel & Bogorad, 1976).

Although these soluble and insoluble RNA polymerases were considered to be distinct for many years, it now seems likely, based on their similar sensitivities and insensitivities to a range of inhibitors, that the soluble and insoluble RNA polymerases represent the same enzyme in its actively transcribing state (insoluble and bound to the genome) and enzyme which is free in the chloroplast stroma (soluble).

1.4.3 RNA polymerase activities in etioplasts and chloroplasts

Two RNA polymerase activities were isolated from mustard (*Sinapis alba*) and were separated on the basis of their different activities in etioplasts and chloroplasts. These RNA polymerase activities have been well characterised. Reiss and Link isolated TACs from both etioplasts and chloroplasts. The enzymes share common polypeptides, but some are enriched in either the chloroplast or the etioplast TAC. In addition the transcriptional activity of the chloroplast TAC is more than ten times higher than the etioplast TAC, although both TACs transcribe plastid rRNA genes and protein coding genes (Reiss & Link, 1985). It was later noted that the etioplast and chloroplast enzymes have distinct biochemical characteristics such as different chromatographic behaviour. The chloroplast enzyme readily dissociates from associated sigma like factors (SLFs) whereas the etioplast enzyme does not,

these differences being due to phosphorylation of the enzymes. Treatment of the chloroplast enzyme with protein kinase converted it to an etioplast-like form and treatment of the etioplast enzyme with phosphatase conferred chloroplast-like properties upon the enzyme. Phosphorylation of both the core enzyme and the SLFs causes maximal reconstitution of the holoenzyme, although phosphorylation of either the core enzyme or the SLFs will cause partial reconstitution. The SLFs also have different properties in etioplasts than in chloroplasts. For example, the SLF^{52} has different ionic strength requirements for optimal activity in etioplasts than in chloroplasts and the etioplast SLF^{29} and chloroplast SLF^{29} have different preferences for the *psb*A promoter. These differences are due to the phosphorylation state of the SLF (Tiller & Link, 1993). A protein kinase which phosphorylates SLFs was later found associated with the chloroplast RNA polymerase in mustard (Baginsky *et al.*, 1997).

Subsequent work identified two RNA polymerases in mustard chloroplasts and etioplasts designated A and B. Both are large multi-subunit enzymes, the A enzyme has a molecular mass of more than 700 kDa and consists of at least 13 subunits whereas the B enzyme has a molecular mass of 420 kDa and consists of only four subunits. The A enzyme is the predominant activity in chloroplasts and the B enzyme is the predominant activity in etioplasts. The two enzyme activities differ in their sensitivities to rifampicin. The A enzyme is resistant to rifampicin, while the B enzyme is inhibited by rifampicin (Pfannschmidt & Link, 1994). The A and B enzymes are capable of transcribing both plastid photosynthetic and housekeeping genes *in vitro*, although rifampicin inhibits transcription of all gene classes by the B enzyme. Both enzymes use the same transcription start sites. Mustard seedlings grown in the presence of rifampicin have reduced transcript levels of the photosynthetic genes *rbc*L and *psb*A but normal transcript levels of the non-photosynthetic genes *rps*16, *trn*G, *rrn* and *rpo*B relative to an untreated control. This suggests that the B enzyme transcribes photosynthetic genes

and that the A enzyme (or another rifampicin resistant RNA polymerase) transcribes housekeeping genes (Pfannschmidt & Link, 1997).

Further analysis of the A enzyme by *N*-terminal sequencing of the composite polypeptides has identified *rpo* gene products in this complex. The A enzyme is structurally related to the B enzyme and can be converted into a rifampicin sensitive enzyme by phosphorylation (Pfannschmidt *et al.*, 2000).

1.4.4 The evidence for a nuclear-encoded polymerase

The mapping and sequencing of the plastid genome of a non-photosynthetic, parasitic flowering plant (*Epifagus virginiana* – beechdrops) has provided some strong evidence for the existence of a nuclear encoded chloroplast RNA polymerase. *E. virginiana* has a greatly reduced plastid genome and all photosynthetic and chlororespiratory genes are absent from the genome but a number of rRNA genes and ribosomal protein genes are present. Transcripts for the genes which have been retained can be detected by northern analysis, so clearly the plastid is transcriptionally active (dePamphilis & Palmer, 1990). Several chloroplast genomes have been completely sequenced and the *rpo* genes encoding PEP have been identified. These genes are absent from the plastid genome of *E. virginiana*. It was concluded that the genome is transcribed by a nuclear-encoded RNA polymerase which is imported into the plastid. The promoters of the *trn*E and *rps*2 genes encoded within the plastid genome of *E. virginiana* differ to the promoters of the equivalent genes in other plants. The promoter region of the *trn*E gene in most plants contains well conserved –35 and –10 motifs which are absent from the promoter region of the *E. virginiana trn*E gene. The *E. virginiana rps*2 gene also lacks prokaryotic consensus sequences in its promoter regions which are present in the maize *rps*2 gene. This suggested that transcriptional activity detectable in *E. virginiana* plastids was not the product of RNA polymerase genes which had been transferred from the plastid to the nucleus (Morden *et al.*, 1991).

This finding was reinforced by subsequent studies on barley (*Hordeum vulgare*) and rye (*Secale cereale*). Transcriptional activity was detected in the white leaves of a mutant of barley (*albostrians*) and in heat-bleached leaves of rye. Both of these strains lack plastid ribosomes and are unable to translate plastid encoded transcripts. In both barley and rye the genes encoding PEP subunits are plastid-encoded, therefore PEP must be absent in these strains. It was concluded again that the transcriptional activity was a result of a nuclear-encoded RNA polymerase. It was determined that this RNA polymerase activity preferentially transcribes housekeeping genes such as *rps*15 and *rpo* genes over photosynthetic genes as transcripts for housekeeping genes accumulate, but transcripts for photosynthetic genes do not. Transcripts for photosynthetic genes such as *rbc*L, *atp*I, *atp*H and *psb*A are detectable at low levels, but transcripts for housekeeping genes such as *rps*15, *rps*2, *rpo*A, *rpo*B and *rpo*C1 are detectable at levels which seem to be enhanced (Hess *et al.*, 1993).

The advent of efficient chloroplast transformation technology in tobacco allowed the targeted deletion of the *rpo*B gene encoding for the β subunit of PEP. The resulting transformants were non-photosynthetic and pigment deficient. The transformants were maintained on sucrose-containing media to compensate for the lack of photosynthesis. Under these conditions the *rpo*BΔ seedlings germinate but grow more slowly than wild-type plants. The plants develop normally and show no differences in organ morphology. From this it was concluded that PEP was not essential for non-photosynthetic plastid functions. The plastids of *rpo*BΔ plants are smaller and rounder than those of wild-type plants, but larger than undifferentiated proplastids. In addition the plastids do not contain stacked thylakoid membranes which are present in photosynthetically active chloroplasts. Northern analysis determined that transcript levels for a number of photosynthetic genes including the *psb*D/C operon, *rbc*L and *psb*A were greatly reduced compared to wild-type levels. In contrast transcript levels for a number of housekeeping genes including *rpl*16

and *16SrDNA* were comparable to wild-type. As many chloroplast encoded genes are arranged as part of operons and co-transcribed with other genes, the transcript levels of the *atp*I gene were studied by northern analysis. The *atp*I gene is part of the *atp*A operon which contains photosynthetic genes and a housekeeping gene (*rps*2). The transcript levels of the *atp*I gene in the *rpo*BΔ plants were found to be similar to wild-type levels.

The 5′ transcript ends of four chloroplast genes (*rbc*L, *16SrDNA*, *psb*A and *psb*D) were mapped in the *rpo*BΔ plants to establish whether the transcripts were initiated at –10 / -35 prokaryotic-like promoters. None of the 5′ ends mapped to these promoter initiation sites so it was concluded that the RNA polymerase activity in the *rpo*BΔ plants was not due to a prokaryotic like polymerase, but the result of a second distinct plastid transcription system. It was suggested that this must be nuclear encoded since the tobacco chloroplast genome has been completely sequenced and none of the unidentified orfs show sequence similarity to RNA polymerases.

In summary the preferential accumulation of transcripts from housekeeping genes in the *rpo*BΔ plants suggests that the nuclear encoded polymerase transcribes the housekeeping genes and that the PEP transcribes photosynthetic genes (Allison *et al.*, 1996).

1.4.5 PEP

The chloroplast PEP is similar to the *E. coli* RNA polymerase in that the core enzyme consists of two α subunits (encoded by the *rpo*A gene) a β subunit (encoded by the *rpo*B gene) and a β′ subunit, which, unlike the *E. coli* RNA polymerase, is split into two polypeptides β′ and β″ (encoded by the *rpo*C1 and *rpo*C2 genes), as it is in cyanobacteria (Hudson *et al.*, 1988) (in *E. coli* the β′ subunit is encoded by one gene, *rpo*C). These genes were identified by the complete sequencing of a large number of plant and algal chloroplast genomes in the 1980's and the *rpo*B, *rpo*C1 and *rpo*C2 genes are present on

all completely sequenced chloroplast genomes to date (except the reduced genome of *E. virginiana* as previously discussed). The *rpo*A gene, however, appears to have been transferred to the nucleus in some species. Further studies showed that these genes are expressed (for example the *rpo*A gene of spinach (Sijben-Muller *et al.*, 1986), but for some time these genes and their gene products were not linked to the transcriptionally active chloroplast RNA polymerases previously isolated by biochemical methods.

This connection was initially made by Lerbs and co-workers who identified subunits which are immunologically related to the *E. coli* α and β or β' subunit in the transcriptionally active fraction from spinach chloroplasts (Lerbs *et al.*, 1985) (Lerbs *et al.*, 1988). Little and Hallick made a more direct link between the putative chloroplast RNA polymerase genes and the transcriptionally active chloroplast fractions. They used antibodies raised to recombinant proteins produced from the spinach *rpo*A, tobacco *rpo*B and *E. gracilis rpo*C genes to immobilise chloroplast DNA-dependent RNA polymerase from spinach, and *E. gracilis*. The antibodies prevent transcription of a spinach chloroplast-encoded tRNA gene by spinach soluble chloroplast RNA polymerase, but have little effect on transcription by the TAC of *E. gracilis* and spinach, suggesting that these genes specify components of the soluble RNA polymerase rather than the TAC (Little & Hallick, 1988).

Antibodies raised against a recombinant portion of the C-terminus of the polypeptide encoded by *rpo*A of maize detect a 39 kDa polypeptide in the chloroplast stromal fraction (Ruf & Kossel, 1988). Antibodies raised against a recombinant protein from the pea *rpo*A gene and antibodies raised against the RNA polymerase core enzyme of *Anabaena* detect a 43 kDa protein in a crude soluble RNA polymerase preparation from pea chloroplasts which is concluded to correspond to the α subunit (Purton & Gray, 1989).

The most completely characterised enzyme in this respect is the soluble chloroplast RNA polymerase of maize. This polymerase, which retains the ability to initiate transcription of the cloned *rbc*L gene, was highly purified and found to contain polypeptides of 180, 120, 85, 78 and 38 kDa. Amino-terminal amino acid sequencing of the 180, 120, 78 and 38 kDa polypeptides found that they correspond to the deduced amino-acid sequences of the maize *rpo*C2, *rpo*B, *rpo*C1 and *rpo*A genes respectively (Hu & Bogorad, 1990), (Hu *et al.*, 1991).

Evidence that the TAC and the soluble RNA polymerase represent the same enzyme in different states comes from the identification of the 38 kDa PEP α subunit by immunological analysis using antibodies raised to a recombinant *rpo*A gene product in both the TAC and soluble RNA polymerase from barley chloroplasts. TACs prepared from barley proplastids, however, contain less immunoreactive protein although the transcriptional activities of the two TACs are comparable (Suck *et al.*, 1996).

Studies on a highly purified soluble RNA polymerase from pea showed, by SDS-PAGE, that this enzyme consisted of ten polypeptides of 120, 110, 95, 83, 81, 75, 54, 51, 42 and 35 kDa and non denaturing PAGE and gel filtration showed that the native enzyme was 620-669 kDa. Antiserum raised to the pea *rpo*A gene cross-reacts with the 42 kDa polypeptide, proving that the eubacterial type polymerase encoded within the chloroplast genome is a component of this soluble transcriptionally active complex. This complex is inhibited by the phytotoxin tagetitoxin, a known inhibitor of chloroplast and *E. coli* RNA polymerase, but is insensitive to rifampicin (Boyer & Hallick, 1998).

Studies on transcription during chloroplast development in barley leaf mesophyll cells have shown that the *rpo*B-*rpo*C1-*rpo*C2 operon is transcribed early in chloroplast development before the maximal transcription of photosynthetic genes such as *rbc*L, *atp*B, *psa*A and *pet*B. In addition *rpo*B

transcript levels decline ten fold relative to *rbc*L during chloroplast maturation (Baumgartner *et al.*, 1993).

Individual targeted disruption of *rpo*A, *rpo*B and *rpo*C1 genes in tobacco results in mutants with an off-white phenotype, which is identical regardless of which gene is disrupted (fig 1.6). The plastids of these mutants have a poorly developed internal membrane system and are photosynthetically deficient. Plastid encoded polypeptides involved in photosynthesis can not be detected by western analysis, but transcription of housekeeping genes such as the *rrn* genes can be detected. The *rpo*$^-$ plastids contain functional ribosomes and are capable of synthesising pigments and lipids, albeit at lower amounts than the wild-type. The *rpo*$^-$ plastids contain processed nuclear-encoded polypeptides which are imported into the thylakoid lumen by the Sec thylakoid translocation machinery, but polypeptides which are imported by the ΔpH thylakoid translocation machinery accumulate as intermediates (Santis-Maciossek *et al.*, 1999).

In *C. reinhardtii*, however, it has been reported that disruption of the genes initially designated as *rpo*B1, *rpo*B2 and *rpo*C2 (Fong & Surzycki, 1992) resulted in heteroplastic transformants, ie the chloroplast maintains a mixture of disrupted and wild-type gene copies suggesting that the genes are essential to this organism (Rochaix, 1995). The genes disrupted in these studies have subsequently been re-sequenced. The *rpo*B1 and *rpo*B2 genes were found to correspond to one gene, *rpo*B, and the *rpo*C2 gene was found to extend much further at the 3′ end (S. Purton, unpublished results).

1.4.6 PEP σ factors

Pre 1996 no genes encoding RNA polymerase σ factors had been cloned from plastid containing species. A large number of higher plant and algal chloroplast genomes had been completely sequenced and no genes encoding PEP σ factors had been identified. In prokaryotes binding of the σ

a b

a) This leaf shows wild-type (green), transformed (white) and mixed (light green) sectors segregating after disruption of one of the subunit genes of PEP in tobacco.
b) Pure white plants can be regenerated from white sectors.

Fig. 1.6 The phenotype of PEP deletion mutants in tobacco (Adapted from http://www.botanik.biologie.uni-muenchen.de/botphys/staff/koop/rpo.html)

subunit to the core RNA polymerase consisting of the α, β and β′ subunits produces a holoenzyme ($\alpha_2\beta\beta'\sigma$). The σ subunit promotes accurate transcription initiation by the RNA polymerase holoenzyme by recognising the prokaryotic promoter consensus sequences (-10 and -35) found upstream of genes. The RNA polymerase can initiate transcription non-specifically without σ factors, but the σ factor increases the affinity of the enzyme for specific promoters. Given that many chloroplast encoded genes contain prokaryotic-like -10 and -35 promoters it seemed likely that σ factors would also confer promoter specificity upon the chloroplast PEP. Bacterial σ factors control differential gene expression during developmental processes such as sporulation. Chloroplast gene expression is differentially regulated in response to changes such as light and plant growth and development so σ factors would be necessary to facilitate this process (Troxler *et al.*, 1994).

In non-photosynthetic prokaryotes the σ^{70} family of σ factors are divided into three groups. Group 1 are the primary σ factors. They are responsible for most RNA synthesis in exponentially growing cells and are essential to cell viability. As all group 1 σ factors recognise closely related promoter sequences they are highly conserved in regions involved in promoter recognition. Group 2 σ factors are similar in terms of amino-acid sequence to group 1 σ factors, but are dispensable for exponential cell growth. Group 3 σ factors are the alternative σ factors and differ from group 1 and 2 σ factors in amino-acid sequence. Alignments of the amino-acid sequences of bacterial σ factors have identified four highly conserved regions. Region 1 is divided into two sub-regions 1.1 and 1.2. Region 1.1 is only present in group 1 σ factors, but several residues in region 1.2 are conserved amongst all σ^{70} factors and are thought to be structurally and functionally important, although no function has been assigned. Region 2 consists of four sub-regions, 2.1, 2.2, 2.3 and 2.4. Region 2.1 is involved in the binding of the σ factor to the core RNA polymerase, and may also, along with region 2.3, be involved in DNA strand opening. This was proposed due to the sequence similarity in these regions

with eukaryotic RNA-binding proteins. There is no experimental evidence to suggest this role for region 2.1, but a mutation in region 2.3 confers characteristics expected of a σ factor defective in DNA strand opening. Region 2.2 is the most conserved sequence segment amongst σ factors, but no function has been assigned to this region. Region 2.4 functions in recognition of the –10 region of the promoter and is highly conserved amongst group 1 σ factors, which recognise the same –10 sequence, but less conserved amongst group 3 σ factors, which recognise different –10 sequences. Region 3 is divided into sub-regions 3.1 and 3.2. Region 3.1 is more conserved and has a helix-turn-helix (HTH) DNA-binding motif. Region 3.2 may also be involved in binding of the σ factor to the core RNA polymerase. Region 4 is involved in recognition of the –35 promoter sequence and is divided into sub-regions 4.1 and 4.2. Region 4.2 also has an HTH DNA-binding motif (Lonetto *et al.*, 1992).

Prior to the molecular cloning of σ factor genes, biochemical work suggested that these factors were present in chloroplasts. Two putative σ factors were isolated from *C. reinhardtii*, one of which had a molecular mass of 51 kDa. It was found that this 51 kDa protein does not initiate transcription by nuclear RNA polymerases I and II but binds to the chloroplast RNA polymerase and *E. coli* RNA polymerase (Surzycki & Shellenbarger, 1976).

Further evidence for the presence of σ factors in chloroplasts came from work on maize. Jolly and Bogorad identified a 27.5 kDa polypeptide (designated S) in purified plastid RNA polymerase which accelerates transcription by the plastid RNA polymerase from circular DNA molecules and promotes preferential transcription from chloroplast DNA sequences which have been cloned into plasmids, although the transcription of some chloroplast fragments are affected by the S polypeptide more than others. The *rbc*L gene is preferentially transcribed by the plastid RNA polymerase when the S polypeptide is added (Jolly & Bogorad, 1980).

This work was reinforced by work on spinach when two polypeptides of 90 kDa and 33 kDa which are immunologically related to *E. coli* σ factors were found to cause correct initiation and transcription of the *rbc*L gene when added to a transcriptionally active purified RNA polymerase along with the α subunit. These σ factors also improve transcription initiation by the *E. coli* RNA polymerase. Interestingly, the σ content varies with leaf age. Old leaves contain none of the 90 kDa polypeptide (Lerbs *et al.*, 1988).

Evidence for the binding of different σ factors to different chloroplast promoters came from work on mustard. The σ-like factors (SLF^{67}, SLF^{52} and SLF^{29}) isolated from chloroplasts fail to bind chloroplast DNA but confer enhanced binding and transcriptional activity upon *E. coli* RNA polymerase transcribing DNA fragments containing a chloroplast promoter. Each SLF binds to the *psb*A, *trn*Q and *rps*16 gene promoters with different efficiencies (Tiller *et al.*, 1991).

Later, Troxler and co-workers identified a number of σ factors from a range of species including plants and algae through immunological studies. Antibodies raised against a σ factor from *Anabaena* were shown to cross react with a 64 kDa polypeptide from purified RNA polymerase preparations of maize, a 64 kDa and a 42 kDa polypeptide from purified RNA polymerase preparations of rice, 82 kDa and 100 kDa polypeptides from purified RNA polymerase preparations of *C. reinhardtii* and a 32 kDa polypeptide from purified RNA polymerase preparations of *Cyanidium caldarium* (Troxler *et al.*, 1994).

The first genes encoding for PEP σ factors were cloned as recently as 1996 from the nuclear genome of the red alga *C. caldarium* by two groups. Tanaka and co-workers cloned a nuclear gene encoding a putative PEP σ factor which they designated *sig*A. The deduced amino-acid sequence of this polypeptide shows strong homology to the σ factors of cyanobacteria. A

recombinant SigA protein was made and shown to increase the promoter specificity of the *E. coli* core enzyme. Antibodies raised to the recombinant protein were used in immunogold electron microscopy to localise the SigA protein to the chloroplast (Tanaka *et al.*, 1996). A cDNA encoding a putative σ factor (*rpo*D) was also cloned from the *C. caldarium* nuclear genome by Liu and Troxler. The deduced amino-acid sequence contains all four conserved regions found in bacterial and cyanobacterial σ factors along with an *N*-terminal extension which could not be aligned with σ factors from bacteria and cyanobacteria but could represent a chloroplast transit sequence. This cDNA was used to identify a gene family of σ factors by Southern analysis of the nuclear genome. *rpo*D transcripts are absent from dark grown cells, but present in illuminated cells. Antibodies raised to the RpoD recombinant protein and antibodies raised against a cyanobacterial σ factor detected a 55 kDa protein in partially purified chloroplast RNA polymerase (Liu & Troxler, 1996).

In 1997 Isono and co-workers cloned three cDNAs designated SIG1, SIG2 and SIG3 from the nuclear genome of *Arabidopsis*. These represented the first σ factors cloned from a multi-cellular eukaryote. Again, the deduced amino-acid sequences of these genes contain domains which are highly conserved amongst bacterial RNA polymerase σ factors. Transcripts for these genes are present in leaves, but not in roots, and are induced in leaves of dark-adapted plants in response to illumination. The *N*-terminal regions of the deduced amino acid sequences of two of these genes (SIG2 and SIG3) are predicted to be chloroplast transit peptides at a probability of more than 90% and were therefore fused to Green Fluorescent Protein (GFP) in order to investigate whether these polypeptides are chloroplast-located. These *N*-terminal sequences do indeed target GFP to the chloroplast providing further evidence that these polypeptides function in chloroplast transcription (Isono *et al.*, 1997).

The initial cloning and characterisation of these σ factor genes has led to the isolation of σ factor genes from a range of plant and algal species. Three more cDNA sequences were cloned from *Arabidopsis* (*sig*A, *sig*B and *sig*C) and analysis of the transcript levels of these genes show similar patterns to those of SIG1, SIG2 and SIG3 (Tanaka *et al.*, 1997). One cDNA encoding a putative σ factor was cloned from rice (*Oryza sativa*) (*Os-sig*A). Again analysis of the transcript levels of this gene showed that the transcript levels were higher in green shoots than in roots or dark-grown etiolated shoots and that transcript levels rise when dark-grown etiolated shoots are illuminated (Tozawa *et al.*, 1998). Work on the transcriptional apparatus of mustard chloroplasts (*Sinapis alba*) identified a cDNA, the deduced amino acid sequence of which is 53 kDa (corresponding to the SLF^{52} isolated from mustard by Tiller and co-workers). The amino acid sequence contains regions conserved amongst bacterial σ factors. There is also an *N*-terminal extension which shows characteristics of chloroplast transit peptides and which directs an *in vitro* synthesised protein into the chloroplast, followed by correct processing. The functional region of the protein was expressed in *E. coli* and purified. This protein confers the *E. coli* RNA polymerase with the ability to initiate transcription from the chloroplast *psb*A promoter. The transcript levels of this gene (SIG1) are higher in light-grown than in dark-grown seedlings (Kestermann *et al.*, 1998). Two cDNAs (*sig*1 and *sig*2) were isolated from maize. Although both genes are present as single copies, Southern analysis under low stringency hybridisation conditions showed that a gene family of σ factors is present in the maize nuclear genome. Broadly, transcript levels show the same patterns as found for the *Arabidopsis* genes, SIG1, SIG2 and SIG3, however transcripts for *sig*1 are detectable in etiolated leaves. Immunological studies using an antibody raised against an epitope within *sig*1 showed that an immunoreactive protein was present in etioplasts. This suggests a role for σ factors in chloroplast development (Tan & Troxler, 1999). A cDNA from wheat (*sig*A) was used to show that the *sig*A transcript was under circadian regulation (Morikawa *et al.*, 1999).

Further σ factor genes were cloned from *C. caldarium.* Oikawa and co-workers cloned two more genes called *sig*B and *sig*C. Recombinant SigA, SigB and SigC proteins show σ factor activity when added to the *E. coli* core enzyme. Northern analysis showed that the transcript levels of *sig*A were not affected by light, but the transcript levels of *sig*B and *sig*C were higher in the light than in the dark (Oikawa *et al.*, 1998). Liu and co-workers cloned a gene which they called *rpo*D2 the transcript of which is also absent from dark grown cells but present in illuminated cells. *rpo*D1 shows a distinct pattern of transcript accumulation to *rpo*D2 suggesting that these σ factors are differentially expressed. This study also found that in all five *C. caldarium* σ factors the amino acid sequences of the –10 and –35 recognition motifs are highly conserved, but not identical, suggesting that they may recognise different promoter elements (Liu *et al.*, 1999). To date no genes encoding for σ factors have been cloned from any green algal species.

Fig 1.7 shows an alignment of the deduced amino acid sequences of genes encoding σ factors from *S. alba*, *C. caldarium* and *E. coli.*

A recent study highlights the extent of evolutionary conservation between plant and bacterial σ factors. The function of the *N*-terminal portions of bacterial σ factors differ to the functions of the *N*-termini of plant σ factors. The C-termini of plants and bacteria however, are similar in that this is the specific region of the protein that functions in promoter recognition. In addition the C-terminus of an *Arabidopsis* σ factor fully complements an *E. coli rpo*D mutant, ie it will recognise all the *E. coli* promoters used by the *E. coli* σ factor (Hakimi *et al.*, 2000).

Of the cloned cDNA sequences of plant σ factors ,only that of mustard sig1 (Kestermann *et al.*, 1998) and three of the *Arabidopsis* genes (Hakimi *et al.*, 2000) have been shown to function as σ factors. Although the other cDNA

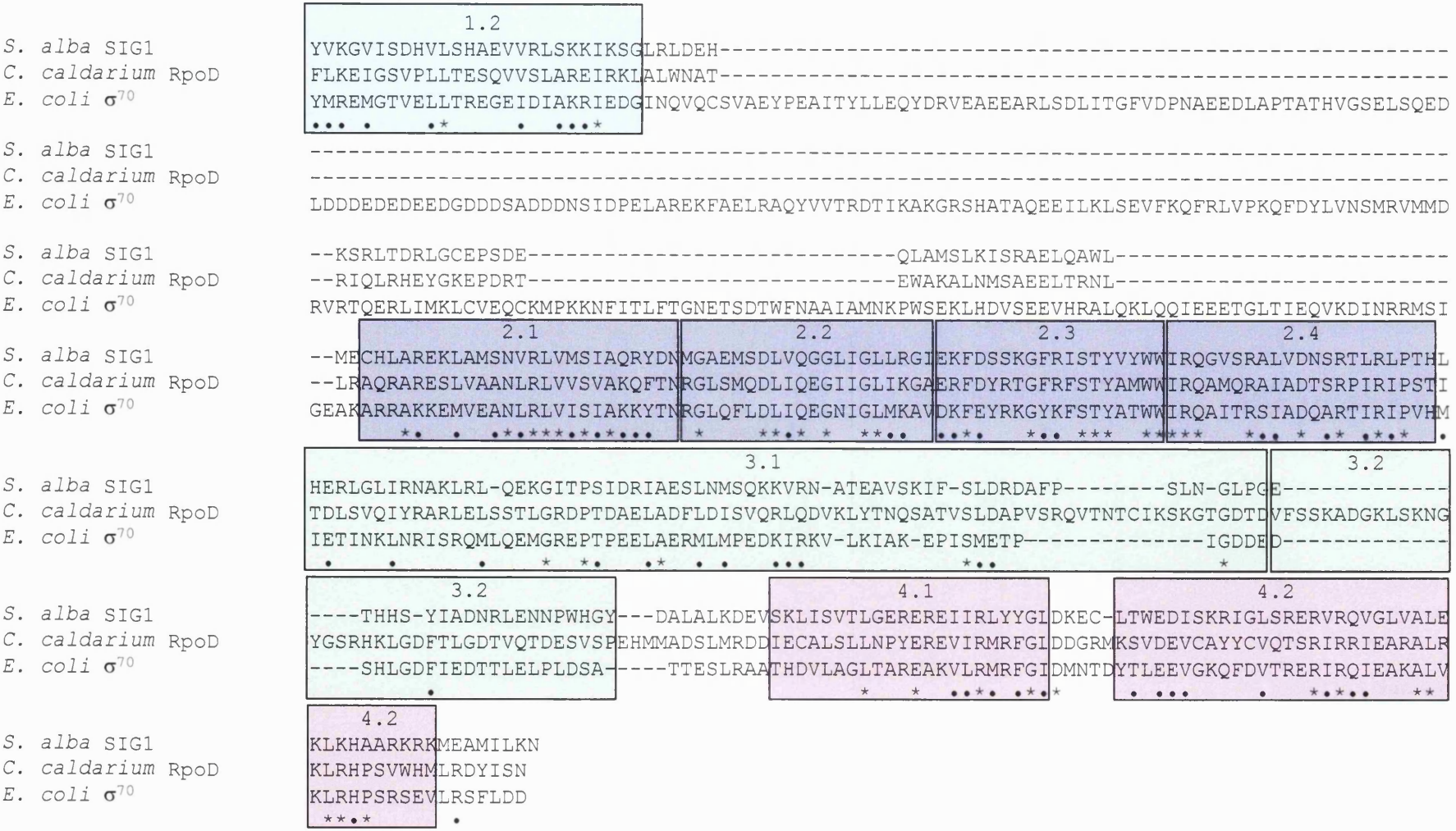


Fig. 1.7 Amino-acid sequence alignment of sigma factors from *S. alba*, *C. caldarium* and *E. coli* (Adapted from Kestermann *et al.*, 1998)

sequences show high degrees of similarity to the bacterial σ factors, and contain the conserved domain regions. The large number of plant σ factors (six from *Arabidopsis*) is surprising given the small size of the plastid genome so it is possible that they perform overlapping functions. The σ factors could also be expressed under different conditions such as developmental specific, tissue-specific or cell-type-specific or in response to light or other conditions. This is the case in bacterial systems and it is true that plastids contain promoters that are regulated in response to all of these conditions. Alternatively it could be that each σ factor is responsible for transcription from a specific promoter allowing the nucleus control over expression of gene families. Plant σ factors have been shown to exhibit promoter specificity (Allison, 2000).

1.4.7 NEP

The first evidence that the nuclear-encoded RNA polymerase was a single subunit phage type RNA polymerase came following work on spinach. Three RNA polymerase activities were biochemically purified from chloroplasts of young spinach leaves, using heparin-sepharose chromatography. The first RNA polymerase activity contained polypeptides immunologically related to the *E. coli* RNA polymerase (ie PEP), but the remaining RNA polymerase activities are associated with a single 110 kDa polypeptide which is unrelated to the *E. coli* RNA polymerase. This purified 110 kDa polypeptide initiates transcription *in vitro* from a bacteriophage T7 promoter, rather than a chloroplast *rbc*L gene promoter, which is efficiently used by *E. coli* RNA polymerase. The 110 kDa polypeptide will transcribe from a fragment containing the *rbc*L promoter in the absence of the T7 promoter to give relatively long transcripts. In addition the T3 and T7 RNA polymerases have a similar molecular mass to the 110 kDa polypeptide, and the 110 kDa polypeptide, in common with the T3 and T7 RNA polymerases, produces more RNA from a supercoiled template than a linear template (Lerbs-Mache, 1993).

This biochemical work has been supported recently by work involving the cloning and characterisation of a nuclear gene which shows similarity to the T7 RNA polymerase and has a chloroplast targeting sequence. The nuclear-encoded mitochondrial RNA polymerase of most eukaryotes is also a single-subunit phage type RNA polymerase. Hedtke and co-workers used the cDNA sequence from the mitochondrial RNA polymerase of *Chenopodium album* to screen an *Arabidopsis* genomic library in an attempt to clone the *Arabidopsis* mitochondrial RNA polymerase. They isolated two clones which represented distinct single copy genes. These genes were termed *rpo*Y and *rpo*Z. The cDNA sequences of these two genes were determined and the predicted amino-acid sequences were 976 (RPOY) and 993 (RPOZ) amino acids long. The molecular masses of the predicted proteins are 111 kDa (RPOY) and 113 kDa (RPOZ). The predicted protein sequences show strong sequence similarity to the yeast nuclear encoded mitochondrial RNA polymerase although the *N*-termini show divergence. The crystal structure of the T7 RNA polymerase has been resolved and catalytically important residues have been identified. All of these residues are present in the predicted amino-acid sequences of RPOY and RPOZ. RPOY contains a transit peptide which can target a GFP fusion protein to the mitochondria and RPOZ contains a transit peptide which can target a GFP fusion protein to the chloroplast (Hedtke *et al.*, 1997). It was suggested by Hedtke and co-workers that the nuclear-encoded chloroplast RNA polymerase may have resulted from a gene duplication event from the mitochondrial RNA polymerase, followed by the acquisition of a chloroplast targeting sequence and re-routing of the protein to the chloroplast.

Genes encoding putative phage like RNA polymerases which are targeted to the chloroplast have subsequently been cloned from maize (Chang *et al.*, 1999) and wheat (Ikeda & Gray, 1999). *In vivo* cellular localisation experiments using GFP showed that the maize protein (encoded by *rpo*Tp) was targeted to the plastid. In addition it was determined that the transcript

levels of *rpo*Tp are higher in leaves than in non-photosynthetic tissues and that transcript levels increased with chloroplast development, then dropped in older tissues, suggesting that RpoTp has a role early in chloroplast development. Transcript levels of *rpo*Tp also increase in response to light. Antibodies raised to the *C*-terminus of RpoTp cross-react with the spinach 110 kDa protein identified by Lerbs-Mache. The cDNA sequence of the wheat gene (wheat-C) encodes a protein with a predicted molecular mass of 107 kDa. The protein has a predicted chloroplast targeting sequence at the *N*-terminus. Phylogenetic analysis using wheat-C shows that this gene clusters closely with the *Arabidopsis rpoZ* gene.

Deletion of *rpo*A, *rpo*B, *rpo*C1 or *rpo*C2 from the chloroplast genome of tobacco yields plants which lack PEP activity, but maintain transcription from NEP promoters. As there are no functional copies of the *rpo* genes in the nucleus this suggests that none of these genes encode essential components of NEP (Serino & Maliga, 1998).

No gene encoding a putative NEP has been cloned from any algal species. Given that it has been suggested that NEP arose from a gene duplication event from the nuclear-encoded mitochondrial RNA polymerase, and that *rpo* gene knockouts in green algae give heteroplastic transformants it is possible that this gene duplication event occurred after the evolution of higher plants (fig 1.8).

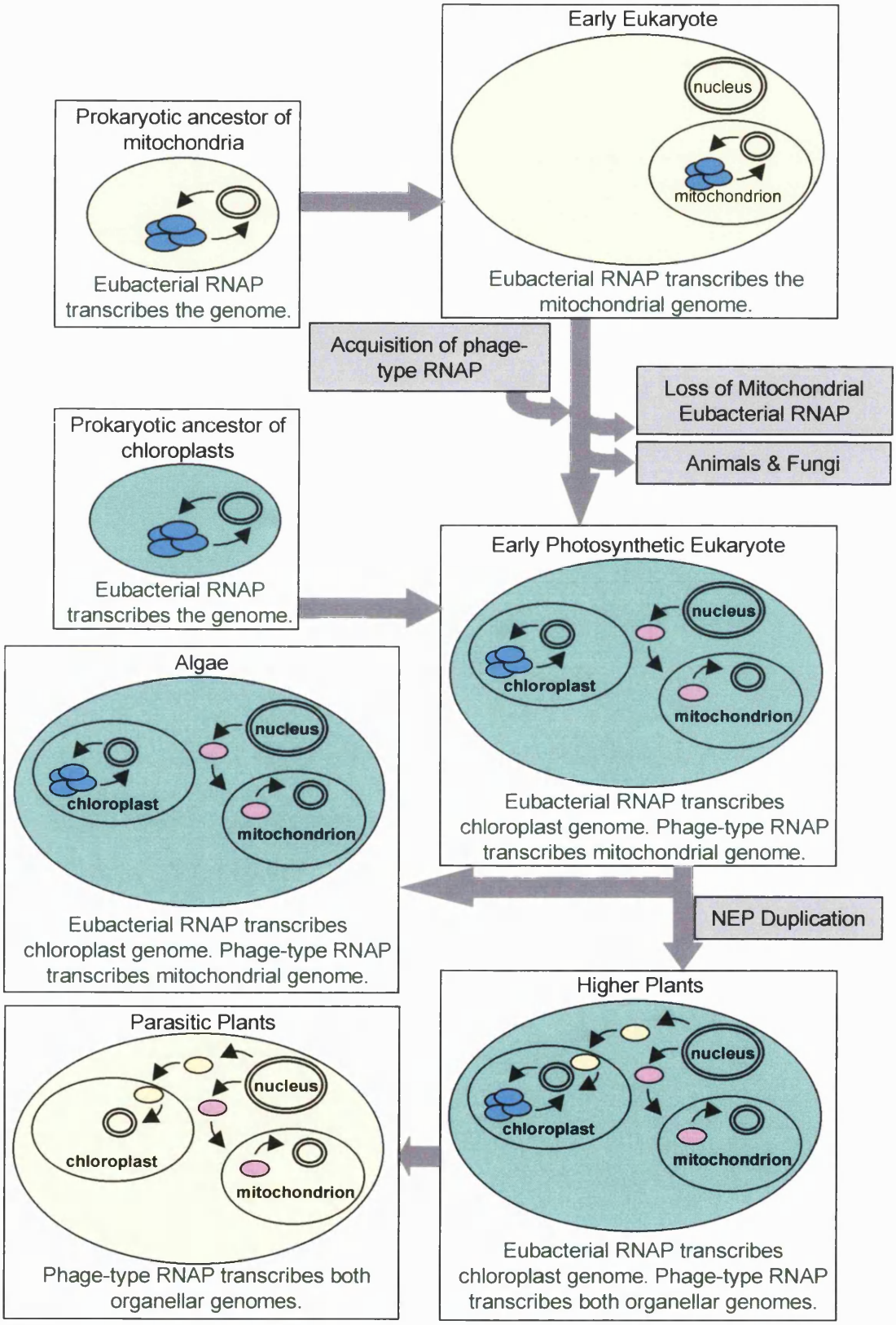


Fig. 1.8 A summary of the model for the evolution of organellar RNA polymerases

Table 1.3 A summary of genes involved in chloroplast transcription

	Chloroplast-Encoded PEP Subunits	Nuclear-Encoded PEP σ Factors	NEP
Higher Plants			
A. thaliana	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2	*Sig*A, *Sig*B, *Sig*C, *Sig*1, *Sig*2, *Sig*3	rpoZ†
Wheat	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2	*Sig*A	Wheat-C†
Maize	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2	*Sig*1, *Sig*2	rpoTp†
Green Algae			
C. reinhardtii	*rpo*B, *rpo*C1, *rpo*C2	*rpo*D	
M. viride	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2		
N. olivacea	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2		
C. vulgaris	*rpo*A*, *rpo*B, *rpo*C1, *rpo*C2		
E. gracilis	*rpo*B, *rpo*C1, *rpo*C2		
Non-Green Algae			
Guillardia theta	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2		
Cyanophora paradoxa	*rpoA, rpoB, rpoC1, rpoC2*		
P. purpurea	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2		
Odontella sinensis	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2		
C. caldarium	*rpoA, rpoB, rpoC1, rpoC2*	*Sig*A, *Sig*B, *Sig*C, *rpo*D1, *rpo*D2	
Protists with Plastids			
P. falciparum	*rpo*B, *rpo*C1, *rpo*C2 (putative *rpo*A in nucleus§)		putative NEP§
T. gondii	*rpo*B, *rpo*C1 *rpo*C2		
Cyanobacteria			
Synechocystis	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2	*rpo*D x4, *rpo*D1, *rpo*E x2, *rpo*F	

*May be non-functional. Lacks C-terminal domain when compared to *E. coli* α-subunit.

§ Identified in EST database

† The nomenclature is still unstandardised for these genes, but all encode for chloroplast-targeted single-subunit phage-type RNA polymerase.

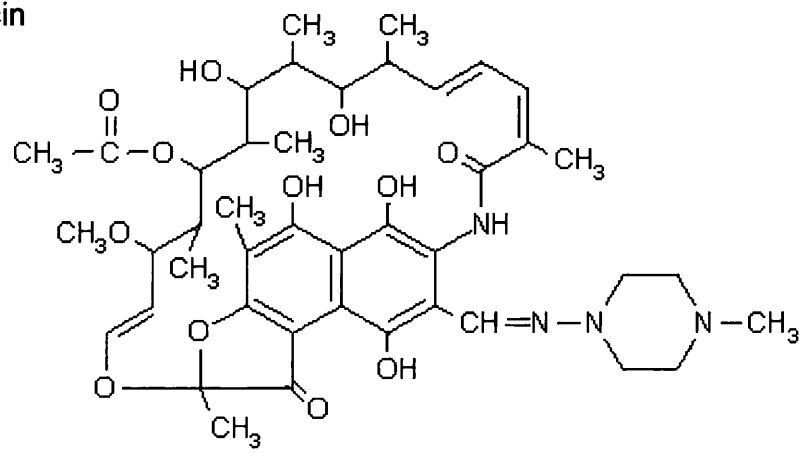
1.4.8 The cascade model of NEP and PEP activity

It has been suggested that NEP and PEP act sequentially to form a cascade during chloroplast development. NEP is active primarily in proplastids and transcribes housekeeping genes. As the plastid develops into a mature chloroplast PEP takes over transcription of housekeeping genes and initiates transcription of photosynthetic genes. This is consistent with observations that photosynthetic genes have PEP promoters and most non-photosynthetic genes have NEP and PEP promoters (discussed in detail later in this section). Also, housekeeping genes are transcribed early in development relative to photosynthetic genes, although transcription from specific NEP or PEP promoters during development has not been shown (Maliga, 1998).

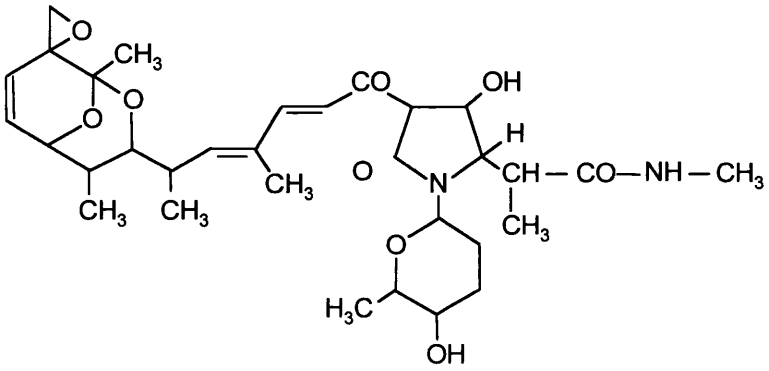
1.4.9 Inhibitors of chloroplast RNA polymerases

As previously discussed the PEP is structurally and functionally related to the *E. coli* RNA polymerase. As such, a number of inhibitors of prokaryotic transcription have been shown to specifically inhibit transcription in higher plant and algal chloroplasts. The structures of some of these are shown in fig. 1.9. Rifampicin is a member of the rifamycin group of antibiotics (Hartmann *et al.*, 1967) and prevents transcription by *E. coli* RNA polymerase and PEP by binding to the β subunit and preventing transcription initiation. Elongation, once initiated, can occur in the presence of rifampicin (fig 1.10). Streptolydigin also inhibits *E. coli* RNA polymerase by inhibiting RNA chain elongation, rather than initiation (Schleif, 1969). Tagetitoxin is a phytotoxin produced by the plant pathogen *Pseudomonas syringae*. This pathogen induces chlorosis in infected plants. This is not due to a direct effect on chlorophyll biosynthesis, but to failed chloroplast development resulting from inhibition of the chloroplast RNA polymerase (Mathews & Durbin, 1994). Tagetitoxin inhibits transcription by chloroplast and *E. coli* RNA polymerase by preventing chain elongation, but has no effect on transcription by bacteriophage T7 or SP6 RNA polymerase (Mathews & Durbin, 1990). Table 1.4 summarises the mechanism of action of these inhibitors.

Rifampicin



Streptolydigin



Actinomycin D

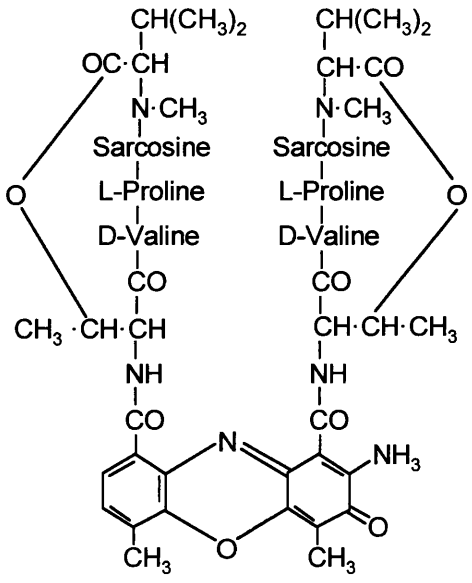


Fig. 1.9 Chemical structures of RNA polymerase inhibitors

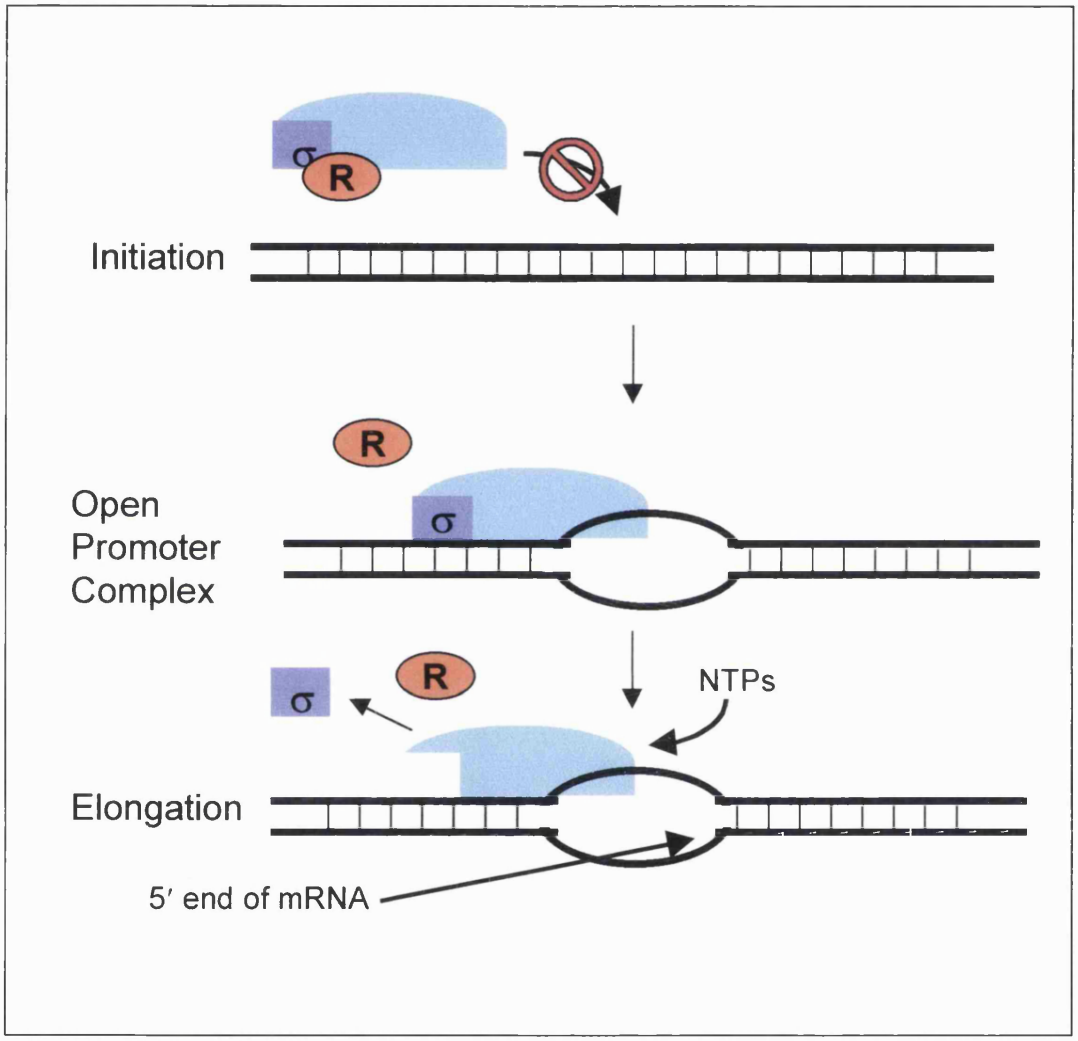


Fig. 1.10 The mechanism of inhibition of PEP by rifamycins. The inhibitor binds to the PEP preventing transcription initiation. During chain elongation the inhibitor is unable to bind to the PEP as the binding site is not exposed.

Table 1.4 Inhibitors of chloroplast and bacterial transcription

Inhibitor	Mechanism of action	Reference
Rifampicin	Chain initiation	(Bottomley, 1970)
Tagetitoxin	Chain elongation	(Mathews & Durbin, 1990)
Streptolydigin	Chain elongation	(Schleif, 1969)
Actinomycin D	Binds to DNA	(Bottomley, 1970)

Early work by Bottomley and co-workers found that neither rifamycin nor streptolydigin have an effect on chloroplast RNA polymerase, leading them to suggest that the chloroplast RNA polymerase does not resemble the bacterial RNA polymerase (Bottomley *et al.*, 1971). These findings are typical of much of the early work on the effect of inhibitors on chloroplast RNA polymerase preparations (sections 1.4.1 and 1.4.2). Tagetitoxin, however, has proved to be a useful tool in the study of higher plant chloroplast RNA polymerases. Tagetitoxin treated wheat leaves lack the large (plastid encoded) and small (nuclear encoded) subunits of ribulose 1,5-bisphosphate carboxylase (RuBPCase) and the corresponding mRNAs as well as 70S ribosomes and their component 16S and 23S rRNAs. The *psb*A transcript is also greatly reduced by tagetitoxin treatment (Lukens *et al.*, 1987). The effect of tagetitoxin on the chloroplast RNA polymerase was confirmed when it was shown to reduce the incorporation of [^{32}P]UTP into RNA by transcriptionally active chloroplast preparations from pea (Mathews & Durbin, 1990). Recently Sakai and co-workers found that tagetitoxin has different effects on transcriptional activities isolated from tobacco chloroplasts and proplastids. Transcription of the *psb*A, *atp*A, *rpo*B, *psa*A, *psa*B, *atp*B, *rbc*L, *pet*B, *rpl*16 and *rrn*23 genes by the chloroplast extract was sensitive to tagetitoxin, but transcription of these genes by the proplastid extract was relatively insensitive (Sakai *et al.*, 1998). This study in particular lends weight to the hypothesis that the nuclear encoded chloroplast RNA polymerase has similar properties to the bacteriophage T7 RNA polymerase, and is active early in chloroplast development. Sakai and co-workers named the chloroplast

transcriptional activity TS-RNA polymerase (tagetitoxin sensitive) and the proplastid transcriptional activity TI-RNA polymerase (tagetitoxin insensitive). They propose that the TS-RNA polymerase corresponds to PEP and that the TI-RNA polymerase corresponds to NEP. The effect of tagetitoxin on transcription in algal chloroplasts has not been reported.

Rifampicin, however, has been used extensively to study chloroplast transcription in both higher plants and algae. Rifampicin inhibits *in vivo* plastid rRNA synthesis in the liverwort *Marchantia polymorpha* (Loiseaux *et al.*, 1975) and, as discussed in section 1.4.3, two distinct RNA polymerases have been isolated from mustard with different sensitivities to rifampicin (Pfannschmidt & Link, 1997).

The work on the effect of rifamycins on algal cells has been carried out mainly on *C. reinhardtii*. Early work by Surzycki found that *C. reinhardtii* cells grown in the presence of rifampin lack 16S and 23S ribosomes, and that *C. reinhardtii* cells will grow heterotrophically on acetate-supplemented medium containing rifampin at 250μg/ml, and phototrophically in minimal media at concentrations of up to 250μg/ml. *C. reinhardtii* cells grown on rifampin for several generations contain chloroplasts, albeit with a large amount of internal disorganisation (Surzycki, 1969).

In a separate study Armstrong and co-workers demonstrated that rifampicin reduces the rate of increase of oxygen evolution in a synchronously growing *C. reinhardtii* culture, prevents cytochrome 553 and 563 production, but has no effect on respiratory capacity, chlorophyll synthesis, phosphoribulokinase, ferredoxin and ferredoxin-NADP reductase production (Armstrong *et al.*, 1971). This study also showed that rifampicin has no effect on the production of RuBP carboxylase, but that the production of this enzyme is inhibited by cycloheximide (an inhibitor of translation from nuclear 80S ribosomes) and spectinomycin (an inhibitor of translation from chloroplast 70S ribosomes).

Given that production of other photosynthetic components such as cytochrome 553 is inhibited by rifampicin and spectinomycin and almost unaffected by cycloheximide, these data could support a dual RNA polymerase theory. Studies by Jennings and Ohad show similar results. Chloramphenicol (an inhibitor of translation from chloroplast 70S ribosomes) had an inhibitory effect on the production of many chloroplast membrane proteins and photosynthetic functions whereas the effect of rifampicin was not as marked. Newly synthesised RNA was detected on chloroplast ribosomes in the presence of rifampicin (Jennings & Ohad, 1972). Rifampicin has been shown to completely inhibit incorporation of ribonucleoside triphosphates (NTPs) into *C. reinhardtii* chloroplast RNA at a concentration of 350μg/ml (Guertin & Bellamere, 1979). More recently a rifampicin resistant chloroplast RNA polymerase has been isolated from *C. reinhardtii* and has been demonstrated to transcribe the chloroplast *trn*E gene, encoding for $tRNA^{Glu}$ *in vitro* (Jahn, 1992).

The effect of rifampicin has also been tested on the giant unicellular green alga *Acetabularia mediterranea (acetabulum)*. Low rifampicin concentrations have no effect on the circadian rhythm of photosynthetic capacity, but do inhibit chloroplast RNA synthesis (Driessche *et al.*, 1970).

1.4.10 Promoters and transcriptional regulation

Gene promoters are an important element in the regulation of gene transcription as they are the DNA regions to which the RNA polymerase binds in order to initiate transcription. The similarity of a large number of chloroplast gene promoters to prokaryotic like promoters has long been noted. Prokaryotic promoters consist of two regions of conserved sequences which are located approximately 10 and 35 bases upstream of the transcription start site and which are optimally separated by 17 bases. A consensus sequence for prokaryotic promoters has been established with the sequence 35 bases from the transcription start site (the "-35" region or the

recognition site) being TTGACA and the sequence 10 bases from the transcription start (the "-10" region or Pribnow box) being TATAAT. Many chloroplast genes also contain these –35 and –10 sequences and many have been shown to be functional, for example the *rbc*L and *atp*B genes of spinach, wheat, maize and pea, the 16S rRNA, $tRNA^{Val}$ and $tRNA^{His}$ genes of maize and the *psb*A gene of mustard. However functional promoters which do not conform to this consensus have also been identified (Kung & Lin, 1985).

A blue light responsive promoter has been identified upstream of the *psb*D-*psb*C genes from barley, wheat, rice, maize, sorghum, tobacco, spinach and pea. The promoter is located within a highly conserved DNA region of 8-25 bp and differs to the prokaryotic like –10 / –35 promoter (Christopher *et al.*, 1992).

Kapoor and co-workers identified an additional functional promoter upstream of the *atp*B/E operon of tobacco which does not show homology to the prokaryotic consensus sequence (CT) which they call non consensus type II (NC-II). Levels of transcripts driven from CT promoters are high in light-grown seedlings, but fall when the plants are transferred to the dark. Levels of transcripts driven from NC-II promoters remain at a constitutive low level in light grown seedlings and that level remains the same when the seedlings are transferred to the dark. The addition of tagetitoxin (a specific PEP inhibitor) reduces the levels of CT transcripts, but has no effect on NC-II transcripts whereas the addition of cycloheximide (an inhibitor of cytoplasmic protein synthesis) reduces levels of both CT and NC-II transcripts. The *rrn*16 and *rpl*32 genes of tobacco also have NC-II promoters (Kapoor *et al.*, 1997).

Hajdukiewicz and co-workers identified promoters for NEP and PEP upstream of tobacco plastid genes. The PEP promoters correspond to the prokaryotic-like –10 and –35 promoters and the NEP promoters are located upstream of the transcription start site, and have a loose consensus of ATAGAATA/GAA. Plastid genes were then assigned to three classes on the basis of their

promoters. Class I genes are mainly photosynthetic and have only PEP promoters. *rbc*L, *psa*A, *psb*A, *psb*B, *psb*D, *psb*E, *pet*B, *ndh*A and *rps*14 are all Class I genes. Class I gene transcripts are high in wild-type plants, but low or absent in *rpo*BΔ plants. Class II genes encode diverse functions and have both PEP and NEP promoters. *rrn*, *atp*B, *atp*I and *clp*P are all Class II genes, and their transcripts are present in both wild-type and *rpo*BΔ plants. No genes encoding for subunits of PSI or PSII are Class II genes. Class III genes have only NEP promoters and are rare, but include *acc*D and *ycf*1. Transcript levels of Class III genes are higher in *rpo*BΔ plants than in wild-type plants (Hajdukiewicz *et al.*, 1997). Studies on the maize *iojap* mutant, which lacks PEP, has also identified NEP promoters which share the consensus of ATAGAATA/GAA (Silhavy & Maliga, 1998). Alignment of NEP promoters identified a CATA or TATA (YATA) sequence, which is essential to the function of the tobacco *rpo*B NEP promoter, and is present in plant mitochondrial promoters (Liere & Maliga, 1999).

NEP promoters have recently been divided into three classes on the basis of their consensus sequences. All but one NEP promoter so far identified have a conserved sequence of 15 nucleotides upstream of the transcription start site. This sequence consists of a highly conserved YRT core motif which is flanked by less conserved AT rich sequences. This is the YRT-box. Class Ia NEP promoters only have this YRT-box. Some NEP promoters also have a conserved sequence of 10 nucleotides upstream of the YRT-box, the GAA-box and these are Class Ib NEP promoters. Class II NEP promoters have a sequence from –5 to +25 nucleotides relative to the transcription start site which is able to initiate transcription (Weihe & Borner, 1999). Fig 1.11 shows a summary of the consensus sequences of chloroplast gene promoters.

Fig. 1.11 (a) PEP promoter consensus sequences (Adapted from Kung & Lin, 1985).

		-35		-10	
E. coli consensus		**TTGACA**		**TATAAT**	
Tobacco	*rbc*L	AAGTAAAAAAGAAAAATTGGG	**TTGCGC** TATATATATGAAAGAGTA	**TACAAT**	AATGATGTATTTGGCAAATC
	*psb*A	ATAGATCTACATACACCTTGG	**TTGACA** CGAGTATATAAGTCATGT	**TATACT**	GTTGAATAAAAAGCCTTCCA
	*atp*B	TCAGGTTCGAATTCCATAGAA	**TAGATA** ATATGGATGGGATTGTC	**TATAAT**	GATAGACAAATGAAAGACTT
	16S rRNA	AGTTGTTCAAGAATAGTGGCG	**TTGAGT** TTCTCGACCCTTTGACT	**TAGGAT**	TAGTCAGTTCTATTTCTCGA
	$tRNA^{Gly}$	TGATTACCACAATTCCCCTGT	**TCGACA** AAAGTTGCATTTGTA	**TACAAT**	AATCGGATTGTA

Fig. 1.11 (b) NEP promoter consensus sequences (Adapted from Weihe & Borner, (1999))

Class Ia				**YRT**
Tobacco	*rpo*B	TATTCAAGCAGGTTGGAATGTGTATTAT	**CAT**	AATAATGG
	*acc*D	ATATAAAGGGGGTTCCAACATATTAATA	**TAT**	AGTGAAGT
	*rps*2	TGGTTATTTGCTTTGGTAATAAAAAGAA	**TAA**	TGAATAGA
Class Ib		**GAA**		
Tobacco	*rpl*32	AATA **GAA** AT---AAGAGTGCCGATCAAG	**TAT**	CCGAAGTC
	*atp*B	AATA **GAA** AATAAAGTTCAGTTCGAATTC	**CAT**	AGAATAGA
	*rrn*16	ATAT **GAA** GCGCATGGATACAAGTTATGC	**CTT**	GGAATGAA

1.5 Aims of this research

The aim of the work presented in this thesis is to characterise the nuclear encoded and chloroplast encoded genes and their gene products involved in the transcriptional apparatus of *C. reinhardtii* chloroplasts.

- To determine whether the *C. reinhardtii* chloroplast gene *rpo*C2, encoding for the β″ subunit of PEP is essential by reverse-genetics using chloroplast transformation.

- To test the sensitivity of the PEP of *C. reinhardtii* to a range of RNA polymerase inhibitors and use these inhibitors to determine whether PEP is essential.

- To use molecular techniques to isolate *C. reinhardtii* nuclear genes involved in chloroplast transcription such as *rpo*A, encoding the α subunit of PEP, genes encoding PEP σ factors and a gene encoding a NEP.

- To further characterise the RpoC2 protein by developing tools to epitope tag this subunit.

Chapter 2 - Materials and Methods

2.1 Chemicals

All chemicals used were of the highest analytical grade available and were purchased from Sigma Chemical Co. (Dorset) unless stated otherwise. All solutions were made using double distilled water (ddH_2O) from a Maxima water filtration unit (Elga, High Wycombe).

2.2 Algal strains

2.2.1 Chlamydomonas *strains*

The strains of *Chlamydomonas* used are listed in Table 2.1 along with the source of this material. The wild-type (WT) *C. reinhardtii* strain CC-1021 mating type+ (mt+) was obtained from Dr E. Harris at the *Chlamydomonas* Genetics Center at Duke University, North Carolina, USA. The ΔpsbH (Bst-opp) mutant, the control strain MluI (O'Connor *et al.*, 1998), the *rpo*C2::*aad*A mutant and the *psb*H-HA mutant were produced and are maintained in this laboratory.

2.2.2 Growth and maintenance of C. reinhardtii *strains*

Media used to culture *C. reinhardtii* are detailed in Table 2.2. *C. reinhardtii* strains were maintained on Tris-acetate phosphate (TAP) medium (Gorman & Levine, 1965) (with appropriate antibiotics where applicable) solidified with 2% (w/v) bacto agar. The cultures were grown under a photon flux of 8μE/m^2/s PAR at 18°C and were streaked to fresh plates every 4-6 weeks. HSM medium was used when photosynthetic growth conditions were required, and the cultures incubated under a higher photon flux of 50μE/m^2/s. Working stocks were re-streaked weekly on solid TAP medium. Liquid cultures were grown, in the appropriate medium, in conical flasks in an

Table 2.1 C. reinhardtii *strains*

***C. reinhardtii* strains**	**Genotype**	**Source**
C. reinhardtii WT CC-1021	wild-type	E. H. Harris - Duke University
*C. reinhardtii cw*10 CC-849	cell wall deficient	E. H. Harris - Duke University
*C. reinhardtii rpo*C2::*aadA*	*aad*A cassette inserted into *rpo*C2 (derived from CC-1021)	S. Nuotio – UCL
C. reinhardtii Δ*psb*H (Bst-opp)	*aad*A cassette inserted into *psb*H (derived from CC-1021)	H. E. O'Connor - UCL
C. reinhardtii MluI	*aad*A cassette inserted into non-coding region downstream of *psb*H (derived from CC-1021)	H. E. O'Connor - UCL
C. reinhardtii Eco10*rrn*S*aadA*	Putative *rrn*S*aad*A cassette inserted into *rpo*C2 (derived from CC-1021)	This thesis
C. reinhardtii Eco10/Tc^R	Putative Tc^R gene inserted into *rpo*C2 (derived from CC-1021)	This thesis
*C. reinhardtii rpo*C2H_6	6x-histidine tagged *rpo*C2 (derived from CC-1021)	This thesis
C. reinhardtii psbH-HA	Putative 3x-HA tagged *psb*H (derived from CC-1021)	N. Ray – UCL

illuminated orbital incubator (New Brunswick Scientific, New Jersey) in the light conditions listed above at 25°C and aerated by shaking at 150 rpm. Starter cultures of 25 ml volume were inoculated from stock cultures and grown to stationary phase (~$2x10^7$ cells/ml) and an appropriate amount of this culture used to inoculate a larger volume of medium. Aseptic techniques were employed throughout.

Table 2.2 C. reinhardtii growth media (Harris, 1989).

For 1 litre	TAP Medium	HSM Medium
H_2O	975ml	925ml
Tris	2.42g	-
4X Beijerink Salts#	25ml	25ml
1M (K)PO_4 pH 7.0*	1ml	-
Trace Elements◊	1ml	1ml
2x PO_4 for HSM†	-	50ml
Glacial Acetic Acid	~1ml to pH7.0	-
Concentrated KOH	-	to pH6.9

4X Beijerink Salts

16g NH_4Cl

2g $CaCl_2$

4g $MgSO_4$

dissolved in 1l ddH_2O

* 1M (K)PO_4

250ml 1M K_2HPO_4

170ml KH_2PO_4 (titrate to pH 7.0)

◊ Trace Elements

I) Dissolve in 550ml ddH_2O in the order indicated below, then heat to 100°C

11.4g H_3BO_4

22g $ZnSO_4.7H_2O$

5.06g $MnCl_2.4H_2O$

4.99g $FeSO_4.7H_2O$

1.61g $CoCl_2.6H_2O$

1.57g $CuSO_4.4H_2O$

1.1g $(NH_4)_6Mo_7O_{24}.4H_2O$

II) Dissolve 50g EDTA.Na_2 in 250ml H_2O by heating and add to the above solution. Reheat to 100°C. Cool to 80-90°C and adjust to pH6.5-6.8 with 20% KOH.

III) Adjust to 1l. Incubate at RT for 2 weeks and allow a rust coloured precipitate to form. The solution will change from green to purple.
IV) Filter through 3 layers Whatman No. 1 paper under suction until the solution is clear.

†2x PO_4 for HSM

14.34g K_2HPO_4

7.26g KH_2PO_4

Media were supplemented with the following antibiotics where necessary:
Spectinomycin (Spc) 100mg/ml in ddH_2O
Streptomycin (Str) 100mg/ml in ddH_2O
Rifampicin (Rif) 100mg/ml in DMSO
Stock solutions dissolved in ddH_2O were filter-sterilised using a syringe and a 0.45μm filter and stored in aliquots at -20°C. Spc and str stock solutions were diluted 1/1000 (final concentration of 100μg/ml) in solid media and 1/2000 (final concentration of 50μg/ml) in liquid culture. Rif stock solution was diluted 1/1000 (final concentration of 100μg/ml) in liquid culture.

2.3 Measurement of cell density

The cell density of *C. reinhardtii* in liquid culture was determined by removing a 1ml sample from an evenly suspended culture. 10μl of tincture of iodine (25mg/ml in ethanol) was added. Duplicate 10μl aliquots of cells killed in this way were then counted using a haemocytometer (Weber Scientific International Ltd., Teddington). The average count was multiplied by 10^4 to give cell density / ml.

2.4 Bacterial strains

2.4.1 Escherichia coli *strains*

The *E. coli* strains used and their genotypes are listed in Table 2.3

Table 2.3 E. coli *strains*

Strain	Genotype	Reference
DH5α	*sup*E44 *hsd*R17 *recA*1 *end*A1 *gyr*A96 *thi-1 rel*A1	(Sambrook *et al.*, 1989)
DK1	Δ(*srl-recA*)306, *araD*139, Δ(*ara, leu*)7697, ΔlacX74, *galU*$^-$, *galK*$^-$, *hsdR*$^-$, strA, *mcrA*$^-$, *mcrB*$^-$	(Kurnit, 1989)
JM109	*recA*1 *sup*E44 *end*A1 *hsd*R17 *gyr*A96 *rel*A1 *thi*Δ(lac-*pro*AB)	(Sambrook *et al.*, 1989)

2.4.2 Growth and maintenance of bacterial strains

Luria-Bertani medium (LB) (Sambrook *et al.*, 1989) was used to culture *E. coli*, unless the medium was supplemented with zeocin, when low-salt LB medium was used. *E. coli* strains were maintained in the short term by growing overnight in liquid media or on media solidified with 2% agar. The appropriate antibiotics were added when necessary, at the concentrations suggested in Sambrook *et al.*, (1989). The cultures were grown at 37°C and then stored at 4°C. *E. coli* cultures were stored long term in the form of glycerol stocks (1ml O/N culture of *E. coli* with 0.5ml LB / 60% glycerol) snap frozen in liquid N_2, then stored at -70°C. Aseptic techniques were employed throughout.

E. coli strains containing the cloning vector pUEX2 and its derivatives were maintained at 30°C due to the presence of a temperature sensitive P_R repressor, encoded by the cI857 gene on the pUEX2 plasmid, which is active at 30°C preventing expression of the LacZ gene (Bressan & Stanley, 1987).

2.5 DNA techniques

2.5.1 Restriction enzyme digestion of DNA

1μg of bacterial plasmid DNA was cut with at least 1 unit of restriction enzyme in the manufacturer's buffer diluted to 1x in ddH_2O, with 1x BSA (0.1 mg / ml) if recommended, in a total volume of 10μl for 1 hour at 37°C. 1μg of *C. reinhardtii* total genomic DNA was cut with an excess of restriction enzyme in 1x buffer and 1x BSA in a total volume of 50μl for 2 hours at 37°C, then an additional 5 units of restriction enzyme was added and the sample incubated for a further 1 hour at 37°C to ensure complete digestion.

2.5.2 Agarose gel electrophoresis of DNA

Restriction enzyme digested DNA fragments or PCR products of greater than 0.5kbp were separated according to size on 1% agarose gels in 1x TAE buffer (40mM Tris acetate, 10mM EDTA.Na_2, pH8) (Sambrook *et al.*, 1989) with Ethidium Bromide at 40μg/ml. The appropriate amount of 6x agarose gel loading buffer (AGLB) (0.1M EDTA.Na_2 pH8.0, 40% Glycerol, 0.01% SDS, 0.01% Bromophenol Blue) was added to the sample which was then subjected to electrophoresis in 1x TAE buffer in a 10cm cooled minigel apparatus (Hoefer, Newcastle-Under-Lyme) at 100V constant voltage for 1-2 hours (plasmid DNA digestions) or in a 30cm maxigel apparatus (Hybaid, Teddington) at a constant voltage of 50V for 14-18 hours (genomic DNA). DNA bands were visualised on a UV transilluminator (UVP, California) and a record kept of each gel using a gel documentation system (UVP, California).

2.5.3 Recovering DNA from agarose gels.

The band containing the required DNA fragment is excised from the gel using a razor blade and the DNA is recovered using the Qiaquick gel purification system (Qiagen, Crawley) according to the manufacturer's instructions. The agarose is solubilised, then the solution is loaded onto an ion exchange

column which is centrifuged in a microcentrifuge (Heraeus, Brentwood) at 13,000 rpm for 1 min, allowing the DNA to bind to the column. The DNA is subjected to ethanol based wash steps, then the DNA is eluted from the column twice with 30µl ddH_2O.

2.5.4 Construction of recombinant DNA plasmids by DNA ligations

The vector is a linearised plasmid which has previously been digested with a restriction enzyme to give compatible ends to the DNA insert. The insert is either the product of a restriction enzyme digestion to give compatible ends to the vector (be they "sticky" or "blunt" ends), or a PCR product (blunt ended unless previously digested with an RE). If both the vector and insert have 5' phosphate groups present they were removed from the vector by treatment with 0.01 units / pmol ends of Calf Intestinal Alkaline Phosphatase (CIAP) (Boehringer Mannheim, Germany). The amount of ends in pmols was deduced by the following equation:

$$\text{pmol ends} = \frac{\text{DNA concentration (µg)}}{\text{kbp size}} \times 3.04$$

The CIAP was added to the restriction enzyme digestion directly following heat inactivation of the enzyme(s), and incubated at 37°C for 1 hour. The vector is then run on a gel and gel purified as previously described.

Vector and insert which had been run on an agarose gel and gel purified as previously described, were used in ligation reactions at a molar ratio of insert : vector of 3:1 for "sticky-ended" ligations or 2:1 for "blunt-ended" ligations. This was determined using the following equation:

$$\frac{\text{ng of vector} \times \text{kbp size of insert}}{\text{kbp size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}} = \text{ng of insert}$$

5 units of T4 DNA ligase (New England Biolabs, Hitchin) were added with ligase buffer diluted to 1x in ddH_2O in a total volume of 10µl. "Blunt-ended" ligations were incubated at 37°C for 1hour and "sticky-ended" ligations were incubated at 16°C overnight.

2.5.5 Preparation of plasmid DNA

Small scale preparations of plasmid DNA were purified from 3ml of overnight cell cultures using a modified version of the plasmid mini preparation method of Sambrook *et al.*, (1989). Larger quantities of plasmid DNA were prepared using Qiagen (Crawley) maxi columns according to the manufacturer's instructions. 100ml of overnight cell cultures were lysed by alkaline lysis and the solution was loaded onto an ion exchange column. The solution flows through by gravity flow and the DNA binds to the column. The column was washed and the DNA eluted using the manufacturer's buffers. This was followed by an isopropanol precipitation step and ethanol washes. DNA from both methods was re-dissolved in ddH_2O.

2.5.6 Preparation of C. reinhardtii *total genomic DNA.*

Total DNA was extracted from *C. reinhardtii* cells using the QIAamp tissue kit (Qiagen, Crawley) adapted for extraction of DNA from plant tissue using the original method for extraction of DNA from *C. reinhardtii* (Rochaix *et al.*, 1988). A 25 ml stationary phase culture was pelleted by centrifugation in a bench top centrifuge (Eppendorf, Germany) at 4,000 rpm for 5 min and the cells resuspended in 0.35ml TEN buffer (50mM EDTA, 20mM Tris-Cl pH8.0, 0.1M NaCl) and transferred to an eppendorf tube. 5µl of Proteinase K at 20 mg/ml and 50µl of 10% SDS were added and the cells incubated at 55°C for 2 hours. The cells were pelleted in a microcentrifuge (Heraeus, Brentwood) at 13,000 rpm for 5 min and the supernatant transferred to a fresh tube. 400µl Buffer AL (supplied by Qiagen) and 40µl RNase A at 20mg/ml were added to the supernatant and the solution vortexed, then incubated at 70°C for 10 min.

420μl of ethanol was added and the solution vortexed again. 620μl of the solution was loaded onto a QIAamp spin column, which is an ion exchange column. The column was centrifuged in a microcentrifuge at 8,000rpm for 1 min allowing the DNA to bind. The flow through was discarded and the rest of the sample loaded onto the column which was then centrifuged as before. The flow through was discarded and the column washed according to the manufacturers instructions. The DNA was eluted twice from the column by incubating the column in 200μl elution buffer (10mM Tris-Cl pH 9.0, pre-heated to 70°C) for 5 min at 70°C, followed by centrifugation as before at 8,000 rpm for 1 min. This gave a final volume of 400μl which was then dried down in a "speedivac" vacuum centrifuge (Uni Equip, Germany) to a volume of 50μl.

2.5.7 Preparation of total genomic DNA from pea

Total cell DNA was extracted from 100mg of pea leaf tissue using the DNeasy plant mini kit (Qiagen, Crawley) according to the manufacturer's instructions. This method involves mechanical cell lysis by grinding 100mg of plant tissue with a mortar and pestle in the presence of liquid N_2. The solution is loaded onto an ion exchange column, and treated as previously described (2.5.6).

2.5.8 Polymerase Chain Reaction (PCR)

PCR was used to amplify DNA from both plasmid and genomic DNA templates prepared as previously described. Approximately 10ng of template was used per reaction. Primers were designed to be largely complementary to the sequences flanking the region of interest and to have similar melting temperatures (T_m). They were supplied by MWG-Biotech (Germany). A final concentration of 1μM of each primer was used in each reaction. 2 units of Vent DNA polymerase (New England Biolabs, Hitchin) were used in a reaction volume of 50μl, containing the reaction buffer diluted to 1x

concentration, all 4 dNTPs at a concentration of 0.2mM each, and a range of $MgSO_4$ concentrations were used from 2mM to 5mM to determine the optimum.

A programmable thermocycler (Techne Progene, Cambridge) was used for the PCR temperature cycles. The template was initially denatured at 95°C for 2.5 min to provide a single stranded template. This was followed by 20-30 cycles of 1 min denaturation at 94°C, 1 min annealing at a temperature of 5°C below the average T_m of the primers and 1-2.5 min extension at 72°C (1 min per 1kb of DNA to be extended). This was followed by an additional 5 min of extension at 72°C.

Alternatively two-step PCR was used for amplifications using primers with long non-homologous tails. Following the initial denaturation step there were 5 cycles of denaturation, annealing and extension at a lower annealing temperature. This was then followed by 20 cycles at the higher annealing temperature.

An approximation of the melting temperature of primers of less than 20 nucleotides can be calculated using the following equation:

$T_m = 2$ (# A+T bases) + 4 (# G+C bases)

For primers of more than 20 nucleotides the following equation is more accurate:

$$T_m = 69.3 + (0.41 \times GC\%) - \left(\frac{650}{\text{length}} \right)$$

The annealing temperature can be calculated using the following equation:

$$\text{Annealing Temperature} = \frac{(T_m \text{ primer \#1} + T_m \text{ primer \#2})}{2} - 5$$

The PCR was analysed by running at least one tenth of the reaction volume on an agarose gel.

2.5.9 DNA sequencing reactions

DNA sequencing was carried out by technical staff at UCL using an ABI Prism 377 Genetic Analyser (Perkin Elmer, California) automated sequencer. The reactions were prepared by cycle sequencing using 500ng of plasmid DNA, 60ng of PCR product of 200-700bp size, 100ng of PCR product of 700-1000bp, 200ng of PCR product of 1.5kbp, or 1μg of PCR product of >3kbp, 10μM primer, AmpliTaq FS and reaction buffer, dNTPs and BigDye labelled terminator ddNTPs (all available from Perkin Elmer, California) in a reaction volume of 20μl. The reactions were ethanol precipitated after 25 cycles of amplification, resuspended in loading buffer and denatured for 3 min at 95°C. Then they were loaded onto the automated sequencer. The returned data was checked manually for base-calling error.

2.6 Transformation techniques

2.6.1 Transformation of E.coli *with plasmids*

Intact plasmids or ligation reactions (carried out as detailed in Section 2.5.4) were used to transform competent *E. coli* cells using a method based on the observations of Mandel and Higa, (1970). The competent cells were made according to a method adapted from Sambrook *et al.*, (1989). The competent cells were used on the same or the next day to transform plasmids, but always used the following day to transform ligation reactions, for maximum efficiency of transformation. To effect transformation, ~100ng of plasmid or all

of a ligation reaction was added to 100µl of competent cells and incubated on ice for 30 min to allow DNA to adhere to the cells. The cells were then heat shocked at 42°C for 90 seconds, to cause DNA uptake, then briefly incubated on ice. 900µl of LB was added and the cells incubated at 37°C for 1 hour to allow expression of antibiotic resistance genes. 200µl aliquots were plated onto media containing the appropriate antibiotic for selection and X-gal (40µg/ml) / IPTG (0.1mM) if blue / white selection was required. The plates were then incubated overnight at 37°C.

Transformation of *E. coli* strains with the pUEX2 plasmid and its derivatives was carried out as described except that heat shock was carried out at 34°C rather than at 42°C and the cells were grown at 30°C rather than at 37°C. This is because the temperature sensitive P_R repressor (encoded by the cI857 gene on the pUEX2 plasmid) is active at 30°C, preventing expression of the LacZ gene. Expression is induced at 42°C and can result in cell death, so heat shock and growth is carried out at lower temperatures (Stanley & Luzio, 1984).

2.6.2 Transformation of E.coli *with cosmids*

The cosmid library containing *C. reinhardtii* total genomic DNA was constructed by Dr Saul Purton (UCL) (Purton & Rochaix, 1995). 100µl of DK1 grown O/N at 37°C in LB was transferred to 10ml of LB supplemented with 0.2% maltose. The cells were grown for 2.5 hours at 37°C. The cells were centrifuged in a bench top centrifuge (Eppendorf, Germany) at 4,000 rpm for 5 min and resuspended in 5ml of 10mM $MgSO_4$. 10µl of phage stock at $8x10^8$ cfu/ml was diluted into 800µl of SM medium (0.1M NaCl, 0.01M $MgSO_4.7H_2O$, 0.05M Tris-Cl pH7.5, 2% gelatin). 100µl of the diluted phage was added to 100µl of DK1 and incubated at 37°C for 20 min. 800µl of LB was added and the sample incubated at 37°C for a further 1 hour. 100µl of the sample was plated onto $LBAmp^{50}$ plates x10 and the plates incubated at 37°C O/N.

2.6.3 C. reinhardtii *chloroplast transformation*

Transformation of *C. reinhardtii* chloroplasts was carried out using the biolistic transformation method. This involves bombarding a lawn of algal cells with microparticles of gold or tungsten coated with the transforming DNA (Boynton *et al.*, 1988). The PDS1000/He helium driven system (Bio-Rad, California) was used to bombard the cells with gold microparticles.

25ml of healthy growing log phase *C. reinhardtii* cells were collected by centrifugation in a benchtop centrifuge (Eppendorf, Germany), resuspended in 5ml TAP agar (0.5%) and poured onto TAP plates containing 100μg/ml spc.

6mg of 1.0μm gold particles were added to 0.5ml 70% ethanol in a microfuge tube. This was vortexed for 2 min, incubated for 15 min, then pelleted by brief centrifugation in a microcentrifuge (Heraeus, Brentwood). The ethanol was removed. The gold was washed with 0.5ml sterile ddH_2O, vortexed, allowed to settle and the ddH_2O removed. This step was repeated twice more, then the gold was resuspended in 100μl sterile 50% glycerol, to give a final concentration of 60mg/ml.

The equipment for the transformation was placed in a laminar flow hood and washed with 70% ethanol before use. The gold particles were thoroughly resuspended by vortexing, then a 25μl aliquot (1.5mg) was removed. 4μl of the transforming plasmid at 1mg/ml, 25μl of 2.5M $CaCl_2$ and 10μl of 0.1M spermidine were added and the tube vortexed vigorously for 3 min. The DNA coated gold microcarriers were pelleted by centrifugation in a microcentrifuge, and the supernatant discarded. 70μl of 70% ethanol was added then removed without disturbing the pellet, then this was repeated with absolute ethanol. The particles were resuspended in 24μl of absolute

ethanol, and 6μl aliquots spread over the centre of the macrocarriers, then left to dry in the laminar flow hood.

The helium gun was assembled according to the manufacturer's instructions, with one agar plate inside and a vacuum was established in the chamber to 27 inches of Hg. The helium flow was switched on and when the pressure behind the rupture disc reached 11,000psi the microparticles were propelled towards the algal lawn. the chamber was vented and the plate removed and placed for incubation at 18°C at 8μE/m^2/s PAR.

2.6.4 Isolation of homoplastic transformants

Transformants appear after about two weeks. When the colonies are at least 1mm in diameter they were picked with a sterile cocktail stick and resuspended in 50μl ddH_2O. This was then spread onto a fresh TAPspc^{100} plate which was incubated under the usual conditions until single colonies appeared. A single colony was again picked and the process repeated at least twice. By this time, the selection pressure for the antibiotic resistance cassette is such that when disrupting a non-essential gene, or modifying a gene in a way which does not affect function, all wild-type copies will have been replaced. Conversely, when disrupting an essential gene, the transfomants will be heteroplastic, and will contain the maximum possible copies of the disrupted / modified gene whilst still retaining cell viability.

Southern blot or PCR analysis was used to confirm the homoplastic / heteroplastic state of the transformant.

2.7 RNA techniques

2.7.1 Electrophoresis of RNA

RNA samples were electrophoresed in 1% agarose gels in 1x MOPS buffer (0.02M MOPS, 8mM Na acetate and 1.5mM EDTA pH7.0 with 2M NaOH) with 1% formaldehyde in a 30cm maxigel apparatus (Hybaid, Teddington) at a constant voltage of 50V for 14-18 hours. The RNA samples were mixed with loading buffer (12μl Ethidium bromide at 12mg/ml, 300μl 10x MOPS buffer, 80μl formaldehyde and 300μl formamide) at a ratio of 6μl loading buffer to 8μl RNA. Prior to loading, the sample was heated to 95°C for 2 min, then cooled on ice. RNA was visualised on a UV transilluminator (UVP, California) and a record kept of each gel using a gel documentation system (UVP, California).

2.7.2 Extraction of total cellular RNA

Total RNA was extracted from *C. reinhardtii* cells using the RNeasy Plant Mini Kit (Qiagen, Crawley) by a method adapted from the manufacturer's instructions. Mid log phase cells were harvested by centrifugation in a benchtop centrifuge (Eppendorf, Germany) at 4,000 rpm for 5 min. The supernatant was removed and the cells resuspended in 450μl of buffer RLT (supplied by Qiagen) and 4.5μl of β-mercaptoethanol was added. The cells were lysed by freezing in liquid N_2 and thawing at room temperature 3 times, then the manufacturer's instructions were followed. The solution is loaded onto an ion exchange column, washed and the RNA eluted with 50μl of ddH_2O treated with DEPC. The yield and purity of the RNA was determined by running an aliquot on an agarose gel, made with TAE buffer as previously described for analysis of DNA.

2.7.3 Extraction of mRNA

mRNA was extracted from *C. reinhardtii* cells using the MicroPoly(A)Pure kit (Ambion, Texas), according to the manufacturer's protocol. $5x10^6$ cells were

lysed by vortexing in the manufacturer's lysis solution, then the cells were centrifuged in a microcentrifuge (Heraeus, Brentwood) at 13,000 rpm for 15 min. The supernatant was incubated with oligo dT cellulose and the polyA tail of mRNA binds to the run of T's attached to the cellulose. The oligo dT cellulose - mRNA was washed with buffers of varying stringencies until the optical density at A_{260} of the flow-through was less than 0.05. The mRNA was stripped from the oligo dT cellulose using the manufacturer's elution buffer. The RNA was ethanol precipitated and resuspended in 10μl of ddH_2O treated with DEPC.

2.7.4 Rapid Amplification of cDNA Ends (RACE)

RACE was carried out according to the method from Bertioli, (1997). 50 units of Avian Myeloblastosis Virus Reverse Transcriptase (AMV-RT) (Boehringer Mannheim, Germany) was used to synthesise a cDNA population from mRNA prepared as previously described, using an oligonucleotide (oligodT) which hybridises to the polyA tail of the mRNA. The reaction was carried out at 42°C for 1 hour. The cDNA population was precipitated with CTAB to remove the enzyme, mRNA, free nucleotides etc. 8.75μl of 1M NaCl and 2.5μl of 10% CTAB was added to the cDNA solution and then centrifuged in a microcentrifuge (Heraeus, Brentwood) at 13,000 rpm for 30 min. The pellet was resuspended in 120μl of 1M NaCl, and incubated for 5 min at RT. 2.7 volumes of ethanol was added and the solution incubated for 10 min at RT. The solution was centrifuged in a microcentrifuge at 13,000 rpm for 10 min and the pellet was washed with 1ml of 70% ethanol. The pellet was briefly dried in a "speedivac" vacuum centrifuge (Uni Equip, Germany) and resuspended in 15μl of DEPC treated ddH_2O.

The second stage of the RACE protocol is a PCR step carried out as previously described using a primer designed to the end of the oligodT primer (RACE 1) and a degenerate primer designed from conserved amino acid sequences within the protein sequence determined from other species. 1μl of

the cDNA was used as a template. Using a thermocycler as previously described the template was denatured at 95°C for 2.5 min to provide a single stranded template. This was followed by 20-30 cycles of 1 min denaturation at 94°C, 1 min annealing at 50°C and 2.5 min extension at 72°C. This was followed by an additional 5 min of extension at 72°C.

The third step is a nested PCR using a primer designed to the end of the oligodT primer (RACE 2), but internal to the RACE 1 primer and a second degenerate primer, designed as before, but to a sequence internal to the original degenerate primer. 1μl of the previous PCR was used as a template and the temperature cycles carried out as before, but at an elevated annealing temperature.

The products of both reactions were analysed by running at least one tenth of the reaction volume on an agarose gel.

2.8 Filter hybridisation of nucleic acids

2.8.1 Southern blotting

DNA was transferred irreversibly to Hybond-N nylon membranes (Amersham Int., Amersham) via the Southern blotting technique (Southern, 1975) as described in Sambrook *et al.*, (1989). The agarose gel was soaked in denaturing buffer (1.5M NaCl, 0.5M NaOH) for 20 min at RT, then briefly washed in dH_2O. Neutralising buffer (1M Tris, pH 8.0, 1.5M NaCl) was added and the gel washed for 15 min at RT, the buffer was replaced with fresh neutralising buffer and incubated for a further 15 min. The gel was again briefly rinsed in dH_2O and the blotting stack assembled. 20x SSC consists of 3M NaCl, 0.3M Na citrate. The DNA was left to transfer to the nylon membrane by capillary action overnight, then the nylon membrane was baked at 80°C for at least 1 hour.

2.8.2 Northern blotting

RNA was transferred irreversibly to Hybond-N nylon membranes (Amersham Int., Amersham) via capillary action as described by Sambrook *et al.*, (1989). The gel was soaked in ddH_2O for 15 min to remove any toxic formamide and formaldehyde, then the blotting stack was assembled as described. The RNA was left to transfer overnight, then the nylon membrane was baked at 80°C for at least 1 hour.

2.8.3 Filter hybridisation of bacterial colonies

Filter hybridisation of bacterial colonies for the purpose of screening a genomic cosmid library was carried out according to a method adapted from Sambrook *et al.*, (1989). Hybond-N nylon membranes (Amersham Int., Amersham) were applied to the plate and left to incubate for two min. The plates and filters were marked to allow subsequent orientation. The filters were transferred, DNA side up, to a tray containing 4 layers of Whatman 3MM paper soaked in denaturing buffer (1.5M NaCl, 0.5M NaOH) and incubated for 5 min. The filters were then transferred to a tray also containing 4 layers of Whatman 3MM paper soaked in neutralising buffer (1M Tris, pH 8.0, 1.5M NaCl) and incubated for 5 min, before being transferred to a final tray containing 4 layers of Whatman 3MM paper soaked in 2X SSC (0.3M NaCl, 0.03M Na citrate). The filters were air dried for 15 min, then baked at 80°C for at least 1 hour.

2.9 Radiolabelling nucleic acid probes

2.9.1 Random primer labelling of DNA probes

DNA probes were radiolabelled with α-^{32}P dCTP (Amersham Int., Amersham) by the random labelling method using the Prime It II Random Primer Labelling kit (Stratagene, California). 25ng of DNA (gel purified restriction enzyme fragment or PCR product) was labelled according to the

manufacturer's instructions. The enzyme was inactivated and the DNA denatured by boiling for 5 min before addition to the hybridisation solution.

2.9.2 *Pulse labelling* in vivo *of RNA transcripts*

The method used to pulse label RNA transcripts in *C. reinhardtii* cells was adapted from those of Gagne & Guertin, (1992), Sakamoto *et al.*, (1993) and Nickelsen *et al.*, (1999). 100ml of *C. reinhardtii* cw10 CC-849 was grown to a cell density of $2x10^6$/ml in TAP. The cells were chilled on ice for 10 min, then pelleted at 5000 rpm for 10 min in a refrigerated superspeed centrifuge (Sorvall, Stevenage) in a GSA rotor. The cells were washed in ~100ml ice cold TAP medium, then pelleted as before and resuspended to $1x10^8$cells/ml in the following chilled buffer: 10mM Tris-Cl pH 8.0, 10mM KCl, 0.5mM EDTA, 1mM β-mercaptoethanol. 1ml of the cells was transferred to a microfuge tube, then subjected to one cycle of freezing in liquid N_2 then thawing at RT. The cells were then pelleted in a microcentrifuge (Heraeus, Brentwood) at 5000 rpm for 5 min and washed in an equal volume of the following chilled buffer : 50mM Tris-Cl pH8.0, 20mM $MgCl_2$, 1mM β-Mercaptoethanol, 50mM NaF, 10% Glycerol. The cells were pelleted and gently resuspended in the previous glycerol buffer to $1.4x10^8$ cells/ml then transferred to a corex tube. The tube was equilibrated at 31-33°C in a shaking incubator at a light intensity of 1000 lux (~20μE). ATP, CTP, and GTP were added to a final concentration of 1mM final and UTP to 0.01mM final, then 0.15 mCi ^{32}P-UTP (Amersham Int., Amersham) was also added. The cells were incubated for 7 min with vigorous shaking then transferred to 2 microfuge tubes. An equal volume of phenol / chloroform was added, and the tubes were vortexed well and centrifuged in a microcentrifuge at 13,000 rpm for 1 min. The supernatant was transferred to a fresh tube and an equal volume of chloroform was added, then the tubes were vortexed well and centrifuged as before. The aqueous phase was transferred to 100μl hybridisation solution and boiled for

5 min. The radiolabelled RNA was then transferred to the tube containing the blot in hybridisation solution.

2.10 Hybridisation of probes

The baked nylon membrane was soaked in 5x SSC at room temperature until thoroughly wet. 40ml of pre-hybridisation solution was prepared containing 5x SSC, 5x Denhardts solution (for 50x Denhardts, 5g Ficoll, 5g BSA, 5g PVP in 500ml dH_2O), 0.5% SDS and 0.2 mg/ml sheared, denatured salmon sperm DNA and pre-heated to 65°C. The wetted nylon membrane was rolled in a filter and then placed in a hybridisation bottle with the pre-hybridisation solution. The bottle and its contents were incubated at 65°C in a hybridisation oven (Hybaid, Teddington) for at least 4 hours.

The pre-hybridisation solution was replaced with 40ml of hybridisation solution and the denatured probe was added. The bottle was incubated at 65°C overnight. The next day the membrane was washed to remove any non-specifically bound probe in the appropriate stringency wash solution, then sealed in a plastic bag and exposed to X-ray film for autoradiography (Kodak, New York) in cassettes containing an intensifying screen (Genetic Research Instruments, Dunmow).

2.11 Protein techniques

2.11.1 Electrophoresis of proteins

Proteins were subjected to SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) using the method of Laemmli (1970). All samples were analysed on acrylamide maxi gels of fixed acrylamide concentrations from 7.5-15%, depending on the size of the protein of interest. The proteins were electrophoresed in a maxigel apparatus (Cambridge Electrophoresis Ltd., Cambridge) at a constant voltage of 50V for 14-18 hours.

Preparation of the samples prior to loading on the gel varied according to the protein, but in general crude cell lysate was prepared by mixing ~5x10^6 cells with an equal volume of 5x Laemmli Gel Loading Buffer (LGLB) (0.3M Tris-Cl (pH 6.8), 50% Glycerol, 10% SDS, 0.01% Bromophenol Blue, 25% β-Mercaptoethanol), boiled for 10 minutes and centrifuged in a microcentrifuge (Heraeus, Brentwood) at 13,000 rpm for 5 min. The supernatant was then loaded onto the gel. Pure protein samples were diluted with an equal volume of LGLB, boiled for 3 min and loaded onto the gel.

To visualise protein bands the gels were stained using Coomassie Blue. The gels were rinsed briefly in destaining solution (40% Methanol, 10% Acetic Acid) then stained with Coomassie Blue solution (Coomassie Brilliant Blue R 1.25g, 50% Methanol, 10% Acetic Acid) for at least an hour. The excess stain was removed by incubating the gel in destaining solution until the bands become clearly visible.

2.11.2 Transfer of proteins to nitrocellulose membranes

Proteins were transferred irreversibly to ECL nitrocellulose membranes (Amersham Int., Amersham) using the western blotting technique as originally described by Towbin and co-workers (Towbin *et al.*, 1979) using a method adapted from Sambrook *et al.*, (1989). The wet transfer method was used in a Trans-Blot™ cell tank (Biorad, California). A sandwich comprising of a nylon wire pad, a sheet of Whatman 3MM paper, nitrocellulose, the SDS-PAGE gel, a second sheet of Whatman 3MM paper and a second pad was clamped together and placed in the tank, immersed in Transfer Buffer (0.1% SDS, 20mM Tris, 150mM Glycine, 20% IMS). The transfer was run at 30V for 14-18 hours in the cold room.

In order to check efficient transfer of proteins to nitrocellulose the blots were stained with Ponceau S (0.2% Ponceau S in 3% TCA) for 5 min, then destained with ddH_2O, until the bands become visible.

2.11.3 Immunological probing of membranes

Immunological probing of the blots was carried out according to the ECL protocol (Amersham Int., Amersham), unless recommended otherwise by the antibody manufacturers. Antibody dilutions varied, and were optimised for each experiment (see Appendix II). All secondary antibodies used were conjugated to alkaline phosphatase and were detected using the ECL chemiluminescence detection kit (Amersham Int., Amersham).

2.11.4 Expression and solubilisation of β-galactosidase 3x-HA epitope fusion protein for SDS-PAGE

5ml of DH5α containing the pUEX2HA plasmid grown O/N in $LBAmp^{50}$ at 30°C was transferred to 100ml of $LBAmp^{50}$ and grown at 30°C for 2 hours. The culture was transferred to 42°C for 2 hours to induce expression of the protein. The cells were pelleted by centrifugation in a refrigerated superspeed centrifuge (Sorvall, Stevenage) at 10,000 rpm for 10 min in a GSA rotor. The cells were resuspended in 5ml of lysozyme at 5mg/ml and incubated on ice for 30 min. 7ml of 0.2% Triton X-100 was added and the cells incubated on ice for a further 10 min. The cultures were centrifuged as before and resuspended in 1ml of 8M Urea, 2% β-mercaptoethanol. An equal volume of 6x LGLB was added and the sample boiled for 3 min. 40μl of this sample was used in SDS-PAGE.

2.11.5 Ni^{2+} column isolation of RpoC2 / PEP

The 6x-histidine tagged RpoC2 protein was isolated from detergent-solubilised cells using a method adapted from Sugiura *et al.*, (1998). Intact

cells are broken by French press, and then separated into soluble and membrane fractions. Both fractions are solubilised with Dodecyl-maltoside, followed by homogenisation, then loaded onto a Ni^{2+}-NTA agarose column in the presence of 25mM Imidazole. The column is washed and the protein is eluted with 200mM Imidazole. 16l of *rpo*C2H_6 culture was grown in TAP to a cell density of 10^7cells / ml, then the cells were harvested using a Millipore Pellicon 0.22μm Harvester (Millipore, Massachucetts), and the concentrated cultures were centrifuged in a refrigerated superspeed centrifuge (Sorvall, Stevenage) at 10,000 rpm for 10 min in a GSA rotor. The pellet was resuspended in a small amount of the supernatant and a small amount of HNG buffer (25mM HEPES (pH 7.0), 100mM NaCl, 10% (w/v) Glycerol) was added. The cells were centrifuged as before and resuspended in 20ml of HNG buffer. 500μl of polyH_6 protease inhibitors (Sigma, Poole) were added, then the cells broken by French pressing, twice, at 6,000lb/in^2. The broken cells were centrifuged as before at 20,000 rpm for 40 min at 4°C in an SS-34 rotor and the pellet resuspended in 20ml HNG buffer. 200μl of inhibitors was added to the soluble (supernatant) and membrane (resuspended pellet) fractions.

The chlorophyll content was measured as an indicator of total protein present and HNG buffer was added to give a final chlorophyll concentration of 1mg/ml. 100μl of inhibitors and Dodecyl-Maltoside to a final concentration of 35mM was added. The samples were homogenised for 30s in the cold room, then centrifuged at 12,000rpm for 10 min in an SS-34 rotor.

The night before use the Ni^{2+}-NTA (Qiagen, Crawley) was equilibrated with water and 200ml equilibrating buffer (HNG-D - HNG + 0.03% (w/v) Dodecyl-Maltoside). The equilibrating buffer was poured off immediately before adding the samples. The soluble and homogenised membrane fractions were poured onto the resin. 100μl of inhibitors and Imidazole to a final concentration of 25mM were added. The solution was mixed slowly in the dark for 30 min, then

poured into the columns and left to settle. After about 15-20 min the column was washed in 200ml wash buffer (HNG-W - HNG + 0.03% (w/v) Dodecyl-Maltoside + 25mM Imidazole), and allowed to run through. The protein was eluted once with 10ml and once with 40ml elution buffer (40mM MES (pH 6.0), 100mM NaCl, 10% (w/v) Glycerol, 0.03% (w/v) Dodecyl-Maltoside, 200mM Imidazole).

The first 10ml of elution were concentrated for SDS-PAGE by TCA precipitation. An equal volume of 10% TCA was added and the sample incubated on ice for 20 min. The sample was centrifuged in a refrigerated superspeed centrifuge at 20,000 rpm for 15 min in an SS-34 rotor. The pellet was washed with 70% ethanol and centrifuged as before. The pellet was resuspended by boiling in a small volume of 1x LGLB.

2.12 Measurement of chlorophyll concentration

Assays of chlorophyll concentration were carried out by adding 100µl of sample to 10 ml of 80% acetone. This suspension was then mixed and filtered through Whatman 3MM paper. The absorbance of duplicate samples was then determined at 645, 647, 652, 653 and 664 nm and an average absorbance calculated for each wavelength. The equation of Porra and co-workers were then used to calculate chlorophyll concentrations (Porra *et al.*, 1989).

2.13 Fluorescence emission spectroscopy

C. reinhardtii cells at the mid-log phase of growth (~$5x10^6$ cells / ml) were pelleted and resuspended to a final cell count of $1x10^7$ cells / ml in growth medium (TAP / HSM). The fluorescence emission of the cells was measured at 77K using a luminescence spectrometer LS-50 (Perkin Elmer, California). The cells were excited at 435nm and fluorescence emission was monitored between 500nm and 750nm.

Chapter 3 – Characterisation of the PEP of *C. reinhardtii*

3.1 Introduction

Although the PEP of higher plants has been very well characterised in recent years (Hess & Borner, 1999) as discussed in 1.4.5, the PEP of *C. reinhardtii* and its corresponding genes have not. The *rpo*B gene encoding the β subunit of PEP was sequenced by Fong and Surzycki (1992). However errors in sequencing lead to the belief that this gene is split in two, *rpo*B1 and *rpo*B2, in *C. reinhardtii*. Errors in the sequencing of the *rpo*C2 gene, encoding the β″ subunit, suggested that the gene was truncated at the 3′ end (Fong & Surzycki, 1992). Goldschmidt-Clermont also identified a gene, *orf*472, which now appears to form part of *rpo*C2 (Goldschmidt-Clermont, 1991). These sequencing errors lead to errors in defining *rpo*C2 and *rpo*B. It was reported that these genes had been disrupted using the *aad*A cassette, and as the resulting transformants were heteroplastic, they appeared to be essential (Rochaix, 1997), (Goldschmidt-Clermont, 1991). However, these interpretations were unreliable due to the mis-sequencing of the genes involved and the selectable marker chosen to disrupt the genes. In light of the evidence from Allison and co-workers that the *rpo*B gene of tobacco was dispensable (Allison *et al.*, 1996) and the cloning of NEP genes from a number of higher plants (Hedtke *et al.*, 1997), (Chang *et al.*, 1999), (Ikeda & Gray, 1999) as discussed in 1.4, it seems necessary to consider the algal situation more closely. The *rpo*C2 gene of *C. reinhardtii* has been re-sequenced in this laboratory and *rpo*C2::*aad*A transformants have been generated but not analysed.

A number of other *C. reinhardtii* chloroplast genes have been identified as being essential on the basis that attempts to knock them out have resulted in heteroplastic transformants. These include orf1995 (*ycf*1), *rps*3, *clp*P and *orf*2971 (Rochaix, 1997). Knocking out genes encoding PEP subunits with an expressed cassette leads to a "catch-22" situation, however. The gene may be essential for transcription of the *aad*A cassette, which is essential for growth on spectinomycin, but the gene is not essential for cell viability *per se*.

Chloroplast RNA polymerase inhibitors have been used to study chloroplast transcription in higher plants and algae but have given very conflicting results as discussed in 1.4.9. In light of the recent advances in the understanding of transcription in higher plants these inhibitors can be used in experiments specifically designed to test the new model.

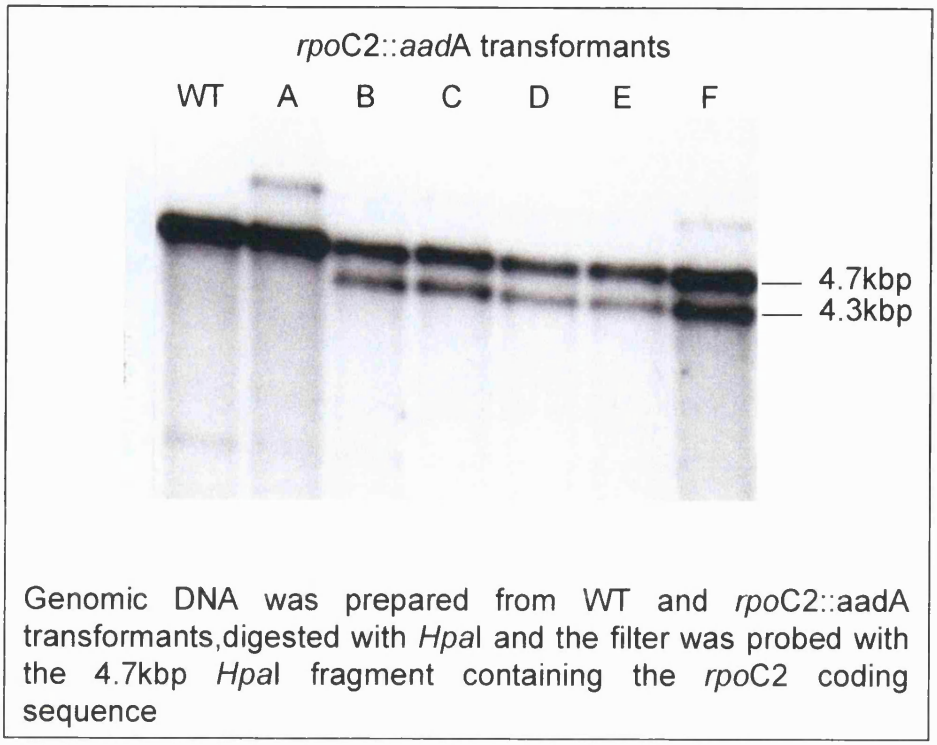
The aims of this chapter are;

- To determine whether the PEP of *C. reinhardtii* is essential by chloroplast reverse genetics using the *rpo*C2 gene.
- To determine a phenotype for the *rpo*C2 knockout transformants.
- To investigate the effects of a range of chloroplast RNA polymerase inhibitors on growth, photosynthesis and transcription in *C. reinhardtii.*

3.2 Results

3.2.1 Southern analysis of rpo*C2::*aad*A transformants*

Southern analysis was carried out in order to determine whether the *rpo*C2::*aad*A transformants generated in this laboratory contained the *rpo*C2 gene disrupted with the *aad*A cassette and to determine the homoplastic / heteroplastic state of the transformants. Total DNA was extracted from six transformants and WT *C. reinhardtii* and digested with *Hpa*I. The digested DNA was run on a 1% agarose gel and transferred to nylon by the Southern blotting technique. A radiolabelled probe was made from the 4.7kbp *Hpa*I fragment from *rpo*C2 and used to probe the blot. Fig. 3.1 shows the resulting autoradiogram. A band of 4.7kbp is seen in the lane containing WT DNA whereas the transformants show a smaller band of 4.3kbp due to the presence of an *Hpa*I site in the *aad*A cassette. This confirms the presence of



Genomic DNA was prepared from WT and *rpo*C2::aadA transformants,digested with *Hpa*I and the filter was probed with the 4.7kbp *Hpa*I fragment containing the *rpo*C2 coding sequence

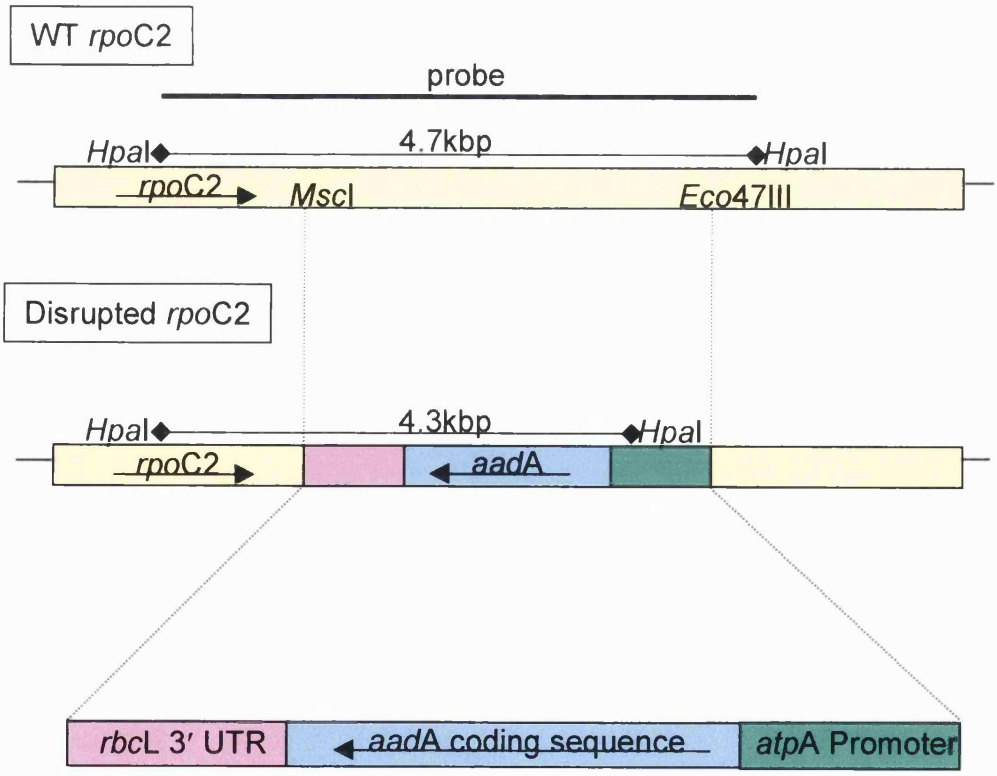
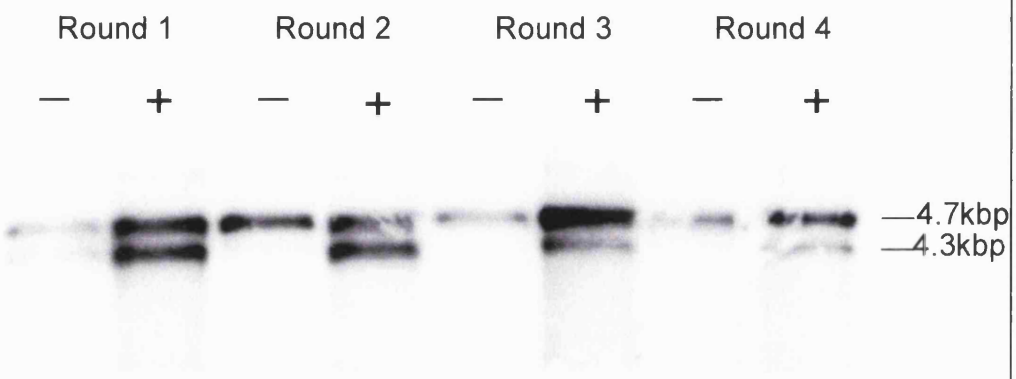


Fig 3.1 Southern analysis of *rpo*C2::*aad*A transformants

the *aad*A cassette in some copies of *rpo*C2. However, the transformants also show the 4.7kbp WT band, so they have maintained WT gene copies. The transformants are heteroplastic suggesting that the *rpo*C2 gene (and hence PEP) is essential to the viability of *C. reinhardtii* cells.

3.2.2 Southern analysis of rpo*C2::*aad*A transformants grown in the absence of spectinomycin*

Heteroplastic gene knockout transformants maintain a balance between the two opposing selective pressures acting upon them; a selective pressure for the (essential) functional gene and a selective pressure for the antibiotic resistance cassette. If the gene is truly essential, when the selective pressure for the antibiotic resistance cassette is removed (ie the transformant is grown on non-selective medium) the remaining selective pressure acting upon the cell (for the functional gene) is strong enough to drive the cell to replace all of the disrupted gene copies with functional gene copies. This should occur within a few rounds of single colony isolation. In order to test this and to gain further insight into the essential nature of *rpo*C2 the *rpo*C2::*aad*A transformant F was taken through four rounds of single colony isolation in the absence of spc and as a control in the presence of spc. DNA was extracted from cells at each round of growth and digested with *Hpa*I as before. The DNA was run on a 1% agarose gel, transferred to nylon via the Southern blotting technique and probed with the 4.7kbp *Hpa*I fragment. The resulting autoradiogram is shown in fig. 3.2. After a single round of growth in the absence of spc the cells still maintain a band of 4.3kbp corresponding to disrupted copies of *rpo*C2. After the second round of growth on non-selective media, however, the 4.3kbp band is not detectable. The cells have replaced most, if not all of the disrupted gene copies with the WT version of the gene. This suggests that the selective pressure for functional *rpo*C2 is very strong, supporting the previous finding that *rpo*C2 is essential to *C. reinhardtii* cell viability.



Genomic DNA was prepared from *rpo*C2::*aad*A transformant F at four rounds of single colony islation in the presence (+) and absence (-) of spectinomycin, digested with *Hpa*I and the filter was probed with the 4.7kbp *Hpa*I fragment containing the *rpo*C2 coding sequence

Fig. 3.2 Southern analysis of *rpo*C2::*aad*A transformant F in the presence and absence of spectinomycin

3.2.3 Spot tests of rpoC2::aadA transformants

In order to compare the growth of the *rpo*C2::*aad*A transformants to WT *C. reinhardtii* and another *aad*A knockout mutant in a range of growth conditions spot tests were carried out. *C. reinhardtii* WT CC-1021, *rpo*C2::*aad*A and *psb*H-null (the *psb*H gene encoding for the 9 kDa polypeptide of PSII is disrupted with the *aad*A cassette (O'Connor *et al.*, 1998)) cells were spotted onto i) TAP plates ii) TAP plates supplemented with spc at 100μg/ml (spc^{100}) and iii) HSM plates. The plates were grown in the light for several days. Fig. 3.3 shows these results. The WT strain grows well on TAP in the light (mixotrophic growth) and on HSM in the light (phototrophic growth) confirming that it is able to photosynthesise. The WT strain does not grow on TAP + spc^{100} in the light, confirming that it is a spc sensitive (spc^{S}) strain. The *psb*H-null strain grows slowly on TAP in the light, when compared to WT. This is because it is non-photosynthetic therefore only capable of heterotrophic growth. As it is non-photosynthetic it is unable to grow on HSM plates in the light (phototrophic growth). It is, however, able to grow slowly on TAP + spc^{100} plates in the light as it contains the *aad*A cassette conferring spc resistance (spc^{R}). The *rpo*C2::*aad*A mutant grows well on TAP in the light, TAP + spc^{100} in the light and HSM in the light. The mutant is spc^{R} and photosynthetically competent. The cell is maintaining enough copies of the WT *rpo*C2 gene to have no effect on photosynthetic ability, but is also maintaining enough copies of *rpo*C2 disrupted with the *aad*A cassette to provide spc^{R}.

The *rpo*C2::*aad*A transformants grow to the same cell density (~$1.3x10^{7}$ cells/ml) and at the same rate as a control strain containing the *aad*A cassette inserted into a non-coding region of the chloroplast genome (MluI) (O'Connor *et al.*, 1998)) when grown in TAP + spc^{50}. The *rpo*C2::*aad*A transformants and the MluI strain grow at a slower rate and to a lower cell density than WT growing in TAP due to the inhibitory effects of spc on chloroplast ribosomes even in the presence of the AAD enzyme.

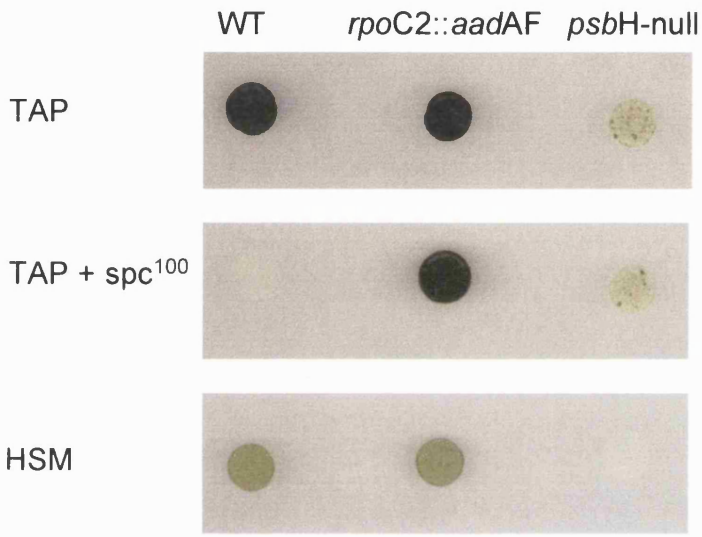


Fig. 3.3 Spot tests of *rpo*C2::*aad*A transformants

3.2.4 Construction of the rpoC2::rrnSaadA knockout gene

In tobacco, disruption of genes encoding PEP subunits with an *aad*A cassette yields homoplastic transformants (Allison *et al.*, 1996), (Santis-Maciossek *et al.*, 1999). However, the higher plant version of the *aad*A cassette is driven from the 16S rRNA gene promoter (Svab & Maliga, 1993). In contrast the *aad*A cassette used in this *C. reinhardtii* disruption is driven from the *atp*A gene promoter (Goldschmidt-Clermont, 1991). Given that PEP transcribes photosynthetic genes in higher plants, it is possible that in *C. reinhardtii* *rpo*C2 is non-essential but that *rpo*C2 cannot be knocked out using an expressed cassette driven from a photosynthetic gene promoter. The *rpo*C2 gene is, therefore, necessary for spc resistance (spc^R), but not for inherent cell viability. In order to eliminate this possibility, the *atp*A promoter of the existing *C. reinhardtii aad*A cassette was replaced with the promoter of the *C. reinhardtii rrn*S gene, encoding for 16S rRNA. This new cassette was used to disrupt *rpo*C2.

The region upstream of the *rrn*S gene that contains the –35 / –10 promoter (Dron *et al.*, 1982) was amplified from the P-228 plasmid (E.H. Harris, Duke University) by PCR using the $rrnSP_R5'$ and the $rrnSP_R3'(2)$ primers. The $rrnSP_R5'$ primer was homologous to the *rrn*S sequence, except that it incorporates an *Eco*RI site to facilitate cloning. The $rrnSP_R3'(2)$ primer was partially homologous to the *rrn*S sequence, but contained a non-homologous tail consisting of a seven nucleotides corresponding to the region upstream of the translation start site of the *C. reinhardtii petA* gene. This region contains the ribosome binding domain, and the *pet*A translation start site (Matsumoto *et al.*, 1991). Since the *rrn*S gene encodes for rRNA, not protein, a ribosome-binding domain must be added to allow translation. The tail also contained the sequence downstream of the translation start site of the existing *aad*A cassette and a *Bsp*HI site which gives a compatible overhang with *Nco*I, to facilitate cloning of the new promoter into the *aad*A cassette. The $rrnSP_R3'(2)$ primer was designed such that the translation start site would be about 250bp

away from the start of transcription site as this is the distance in the *pet*A gene, setting the new promoter in a translational context. The PCR product was digested with *Eco*RI and *Bsp*HI and cloned into the *Eco*RI-*Nco*I sites of the pUC*atp*X*aad*A plasmid. This created the pUC*rrn*S*aad*A plasmid in which the *aad*A cassette is driven from the *rrn*S promoter. The pUC*rrn*S*aad*A plasmid was sequenced with the rrnSPR$_5$′ primer to confirm the presence of the –35 and –10 sequences, the transcription start site, the translation start site and the *Bsp*HI site. The sequence generated is shown in fig. 3.4.

The pUC*rrn*S*aad*A plasmid confers spc^R in *E. coli* at very high levels, comparable to the levels conferred by the pUC*atp*X*aad*A plasmid. Both confer spc^R up to a spc concentration of 500μg/ml.

The *rrn*S*aad*A cassette was amplified by PCR using the rrnSP$_R$5′ primer and the M13 forward primer from the pUC*rrn*S*aad*A plasmid. The PCR product was blunt cloned into the *Msc*I-*Eco*47III sites of the pSKEco10 plasmid such that it is in the opposite orientation to the *rpo*C2 gene. Digestion of the pSKEco10 plasmid with *Msc*I and *Eco*47III drops out a 1.8kbp fragment from the *rpo*C2 gene that is replaced with the *aad*A cassette. This creates the pSKEco10-*rrn*S*aad*A plasmid which contains a deleted / disrupted copy of *rpo*C2.

3.2.5 Transformation of C. reinhardtii *with the pSKEco10-*rrnS*aad*A *plasmid*

The pSKEco10-*rrn*S*aad*A plasmid was used to transform WT *C. reinhardtii* CC-1021 cells using the biolistic method. Selection was on TAP plates supplemented with spc at 100μg/ml (spc^{100}). The resulting transformants were taken through three rounds of single colony isolation on TAP + spc^{100} plates.

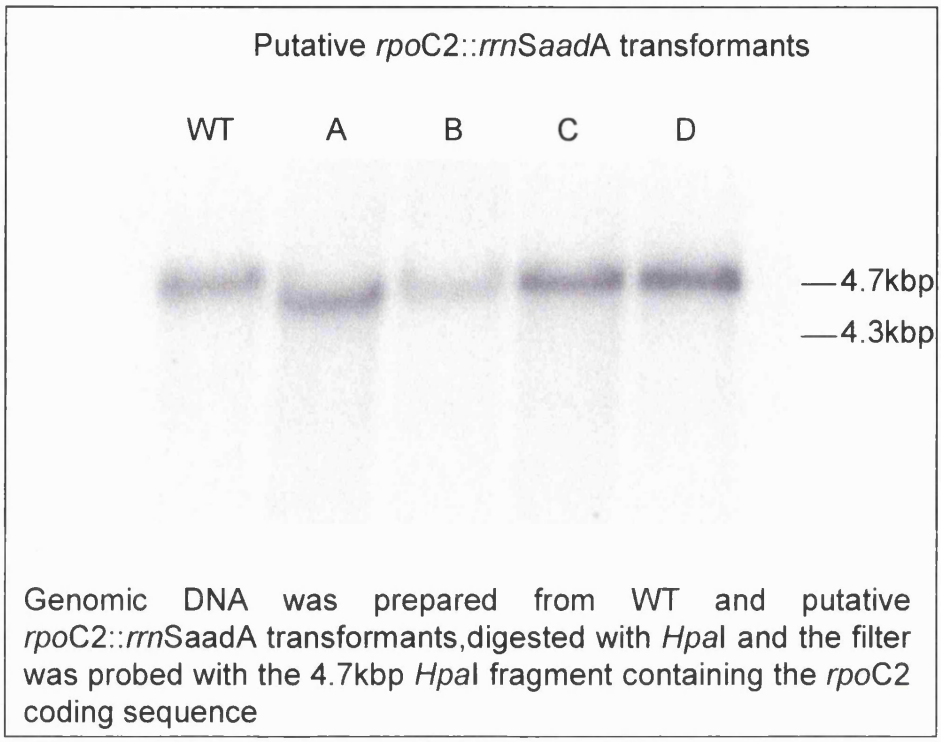
Fig. 3.4 Sequence of the promoter of the *rrn*S*aad*A cassette

3.2.6 *Southern analysis of the putative* rpo*C2::*rrn*Saad*A *transformants*

Southern analysis was carried out on the putative *rpo*C2::*rrn*Saad*A transformants in order to verify that they contained copies of *rpo*C2 that had been disrupted with the *rrn*Saad*A* cassette. Total DNA was extracted from four transformants and WT *C. reinhardtii* and digested with *Hpa*I and *Mlu*I. The digested DNA was run on a 1% agarose gel and transferred to nylon by the Southern blotting technique. A radiolabelled probe was made from the 4.7kbp *Hpa*I fragment from *rpo*C2 and used to probe the blot. Fig. 3.5 shows the resulting autoradiogram. A band of 4.7kbp is seen in the lane containing WT DNA. The presence of an *Mlu*I site within the *rrn*S promoter should yield a smaller band of 4.3kbp if the strains were true transformants. In the lanes containing transformant DNA a band of 4.7kbp corresponding to WT *rpo*C2 is also present, but a band of 4.3kbp corresponding to disrupted *rpo*C2 is absent. It appears that the *rrn*Saad*A* cassette is absent from these transformants and the spc^R must be due to a spontaneous resistance mutation.

3.2.7 *Spot tests of the putative* rpo*C2::*rrn*Saad*A *transformants*

In order to verify that the putative *rpo*C2::*rrn*Saad*A* transformants were in fact the result of a spontaneous resistance mutation and that the *rrn*Saad*A* cassette is absent, WT *C. reinhardtii* and transformant cells were spotted onto i) TAP plates ii) TAP plates supplemented with spc at 100μg/ml (spc^{100}) and iii) TAP plates supplemented with streptomycin at 100μg/ml (str^{100}) and the plates were grown in the light for several days. The *rrn*Saad*A* cassette confers resistance to both spc (spc^R) and str (str^R), however a spontaneous resistance mutation would only confer spc^R. Mutations within the *rrn*S gene which confer spc^R and str^R map to distinct regions so they would be unlikely to spontaneously arise simultaneously (Harris, 1989). Fig. 3.6 shows the results. WT *C. reinhardtii* is sensitive to both spc (spc^S) and str (str^S) and, therefore, only grows on TAP in the light. The putative transformants grow on



Genomic DNA was prepared from WT and putative *rpo*C2::*rrn*SaadA transformants,digested with *Hpa*I and the filter was probed with the 4.7kbp *Hpa*I fragment containing the *rpo*C2 coding sequence

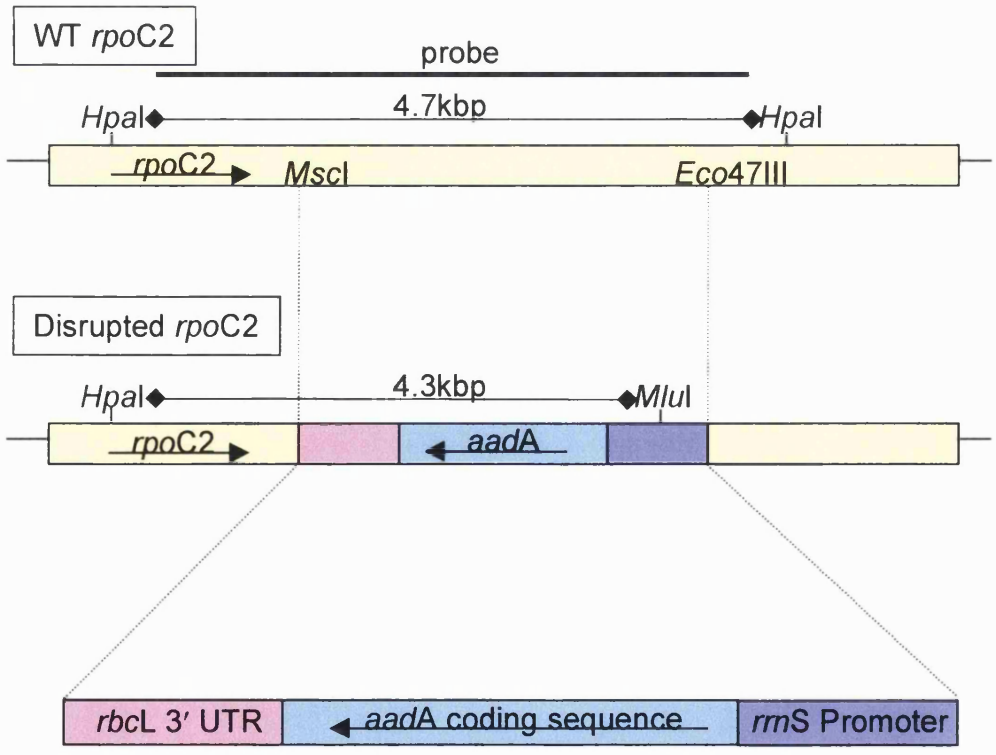


Fig. 3.5 Southern analysis of putative *rpo*C2::*rrn*S*aad*A transformants

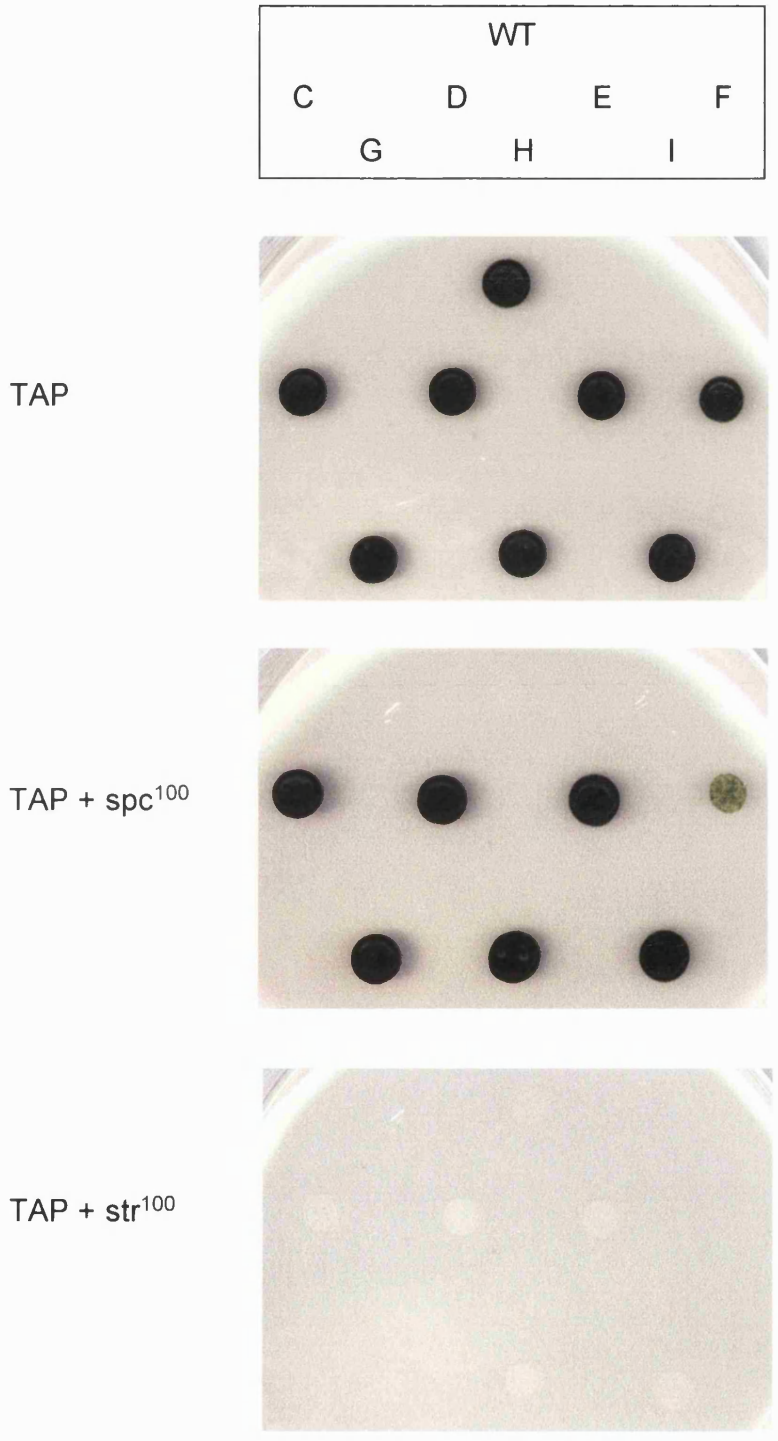


Fig. 3.6 Spot tests of putative *rpo*C2::*rrn*S*aad*A transformants

TAP and TAP + spc^{100} in the light, but do not grow on TAP + str^{100} in the light. The putative *rpo*C2::*rrn*S*aad*A transformants are the result of a spontaneous resistance mutation.

3.2.8 Construction of the rpo*C2::Tc*^{R} *knockout gene*

Non-essential chloroplast genes can routinely be knocked out by disruption with a non-expressed piece of stuffer DNA (Erickson, 1996). Cells are co-transformed with the knockout construct along with a variant of the chloroplast 16S rRNA gene which confers spc^R to allow for selection (fig. 3.7). This approach has been used to successfully disrupt the *atp*B gene (Kindle *et al.*, 1991) and the *rbc*L gene (Newman *et al.*, 1991). In order to investigate the essential nature of *rpo*C2 further this approach was adopted. The *rpo*C2 gene was disrupted with the tetracycline resistance gene (Tc^R) of the cloning vector pBR322.

The Tc^R gene was amplified by PCR from the cloning vector pBR322 using the Tc5′ and Tc3′ primers. The pSKEco10 plasmid containing the *rpo*C2 gene was digested with *Msc*I and *Eco*47III to remove a 1.8kbp fragment. The 1.3kbp Tc^R gene was then blunt cloned into these *Msc*I – *Eco*47III sites in the opposite orientation to *rpo*C2. The resulting pSKEco10-Tc^R plasmid was selected on Tc containing media.

3.2.9 Co-transformation of C. reinhardtii *with the pSKEco10-Tc*^{R} *and P-228 plasmids*

The pSKEco10-Tc^R and P-228 plasmids were used to transform WT *C. reinhardtii* CC-1021 cells plated onto TAP plates supplemented with spc at 100μg/ml (spc^{100}) using the biolistic transformation method. A ratio of pSKEco10-Tc^R : P-228 of 3:1 was used. The P-228 plasmid contains the 16S rRNA gene with an antibiotic resistance mutation conferring resistance to spc

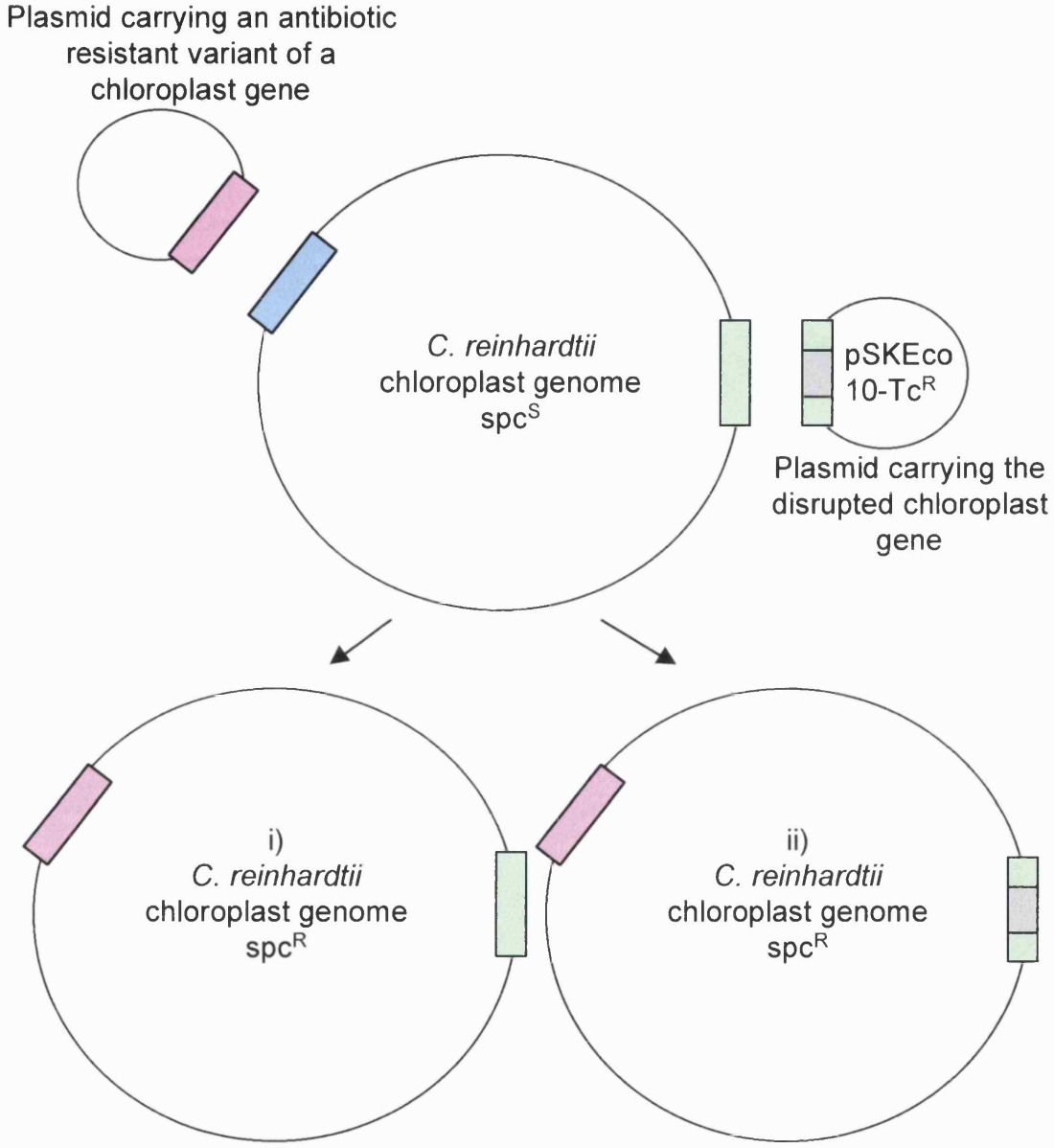


Fig. 3.7 A co-transformation of the chloroplast genome with two plasmids, one conferring resistance to spc and the other carrying the disrupted gene can have two outcomes: i) The antibiotic resistance gene integrates, but the disrupted gene does not; ii) Both changes integrate into the genome

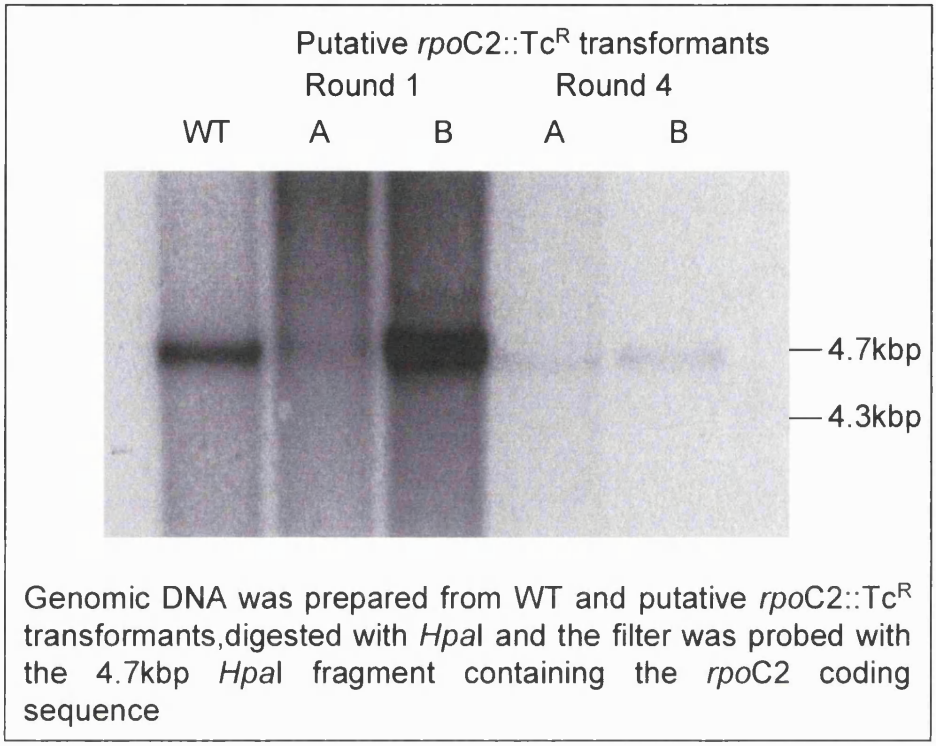
(Newman *et al.*, 1991). The resulting transformants were taken through three rounds of single colony isolation on TAP + spc^{100} plates.

3.2.10 Southern analysis of the putative rpo*C2::Tc*^{R} *transformants*

Southern analysis was carried out on the putative *rpo*C2::Tc^R transformants in order to verify that they contained copies of *rpo*C2 that had been disrupted with the Tc^R gene. Total DNA was extracted from two putative transformants at the first and fourth rounds of single colony isolation and WT *C. reinhardtii* and digested with *Hpa*I. The digested DNA was run on a 1% agarose gel and transferred to nylon by the Southern blotting technique. A radiolabelled probe was made from the 4.7kbp *Hpa*I fragment as before and used to probe the blot. Fig. 3.8 shows the resulting autoradiogram. A band of 4.7kbp is seen in the lane containing WT DNA. The Tc^R gene is 1.3kbp, 0.5kbp smaller than the 1.8kbp removed from *rpo*C2 by digestion with *Msc*I and *Eco*47III. A smaller band of 4.2kbp is expected in the transformants, allowing the presence of the Tc^R gene in *rpo*C2 to be detected. In the lanes containing putative transformant DNA a band of 4.7kbp corresponding to WT *rpo*C2 is also present, but a band of 4.2kbp corresponding to disrupted *rpo*C2 is not detected, even after a single round of growth. It appears that the Tc^R gene is absent from these transformants. This suggests that in the absence of a selective pressure *rpo*C2 will not tolerate a gene disruption, lending further support to the evidence that *rpo*C2 is essential.

3.2.11 Northern analysis of rpo*C2::*aad*A transformants*

In order to study the effect of reduced levels of the *rpo*C2 gene on the cells phenotype northern analysis was performed. Total RNA was extracted from WT CC-1021 grown in TAP, the *rpo*C2::*aad*A transformant F grown in TAP + spc^{50} and the MluI strain grown in TAP + spc^{50}. The RNA was run on a 1% RNA gel and transferred to nylon by the northern blotting technique. A radiolabelled probe was made from the 0.6kbp *Hind*III fragment from the



Genomic DNA was prepared from WT and putative *rpo*C2::Tc^R transformants,digested with *Hpa*I and the filter was probed with the 4.7kbp *Hpa*I fragment containing the *rpo*C2 coding sequence

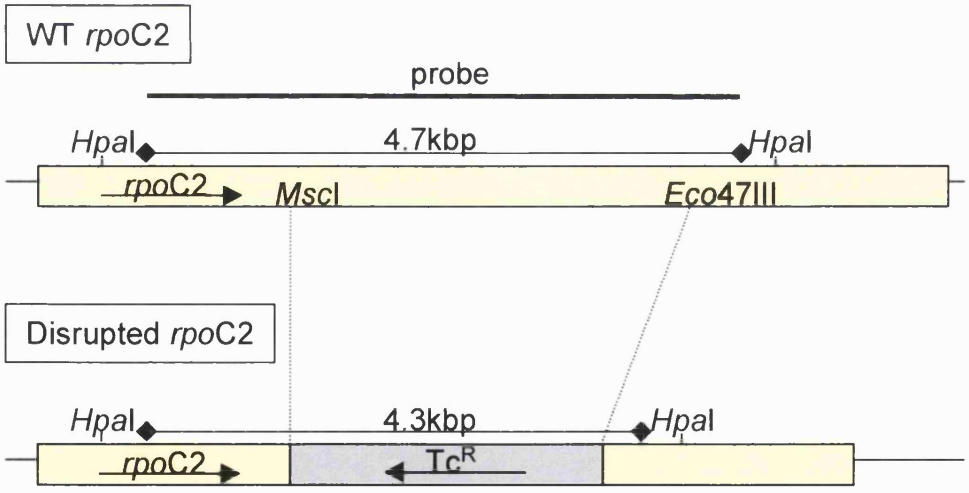
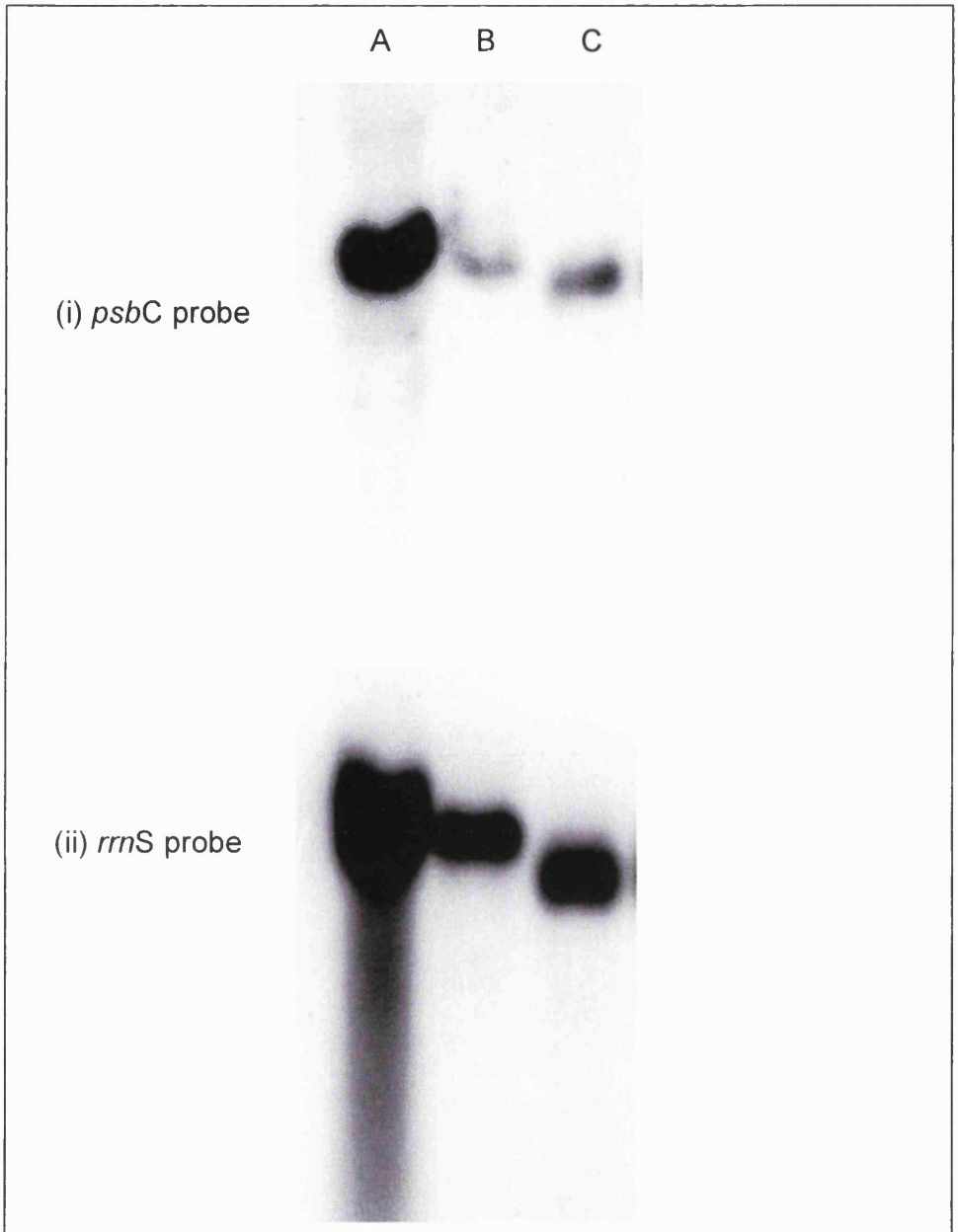


Fig. 3.8 Southern analysis of putative *rpo*C2::Tc^R transformants

*psb*C gene and used to probe the blot. The blot was stripped with boiling 0.1% SDS and reprobed with a radiolabelled probe made from the 4.8kbp *Bam*HI / *Hind*III fragment from the P-228 plasmid containing the *rrn*S gene. The resulting autoradiograms are shown in fig. 3.9. The intensity of the bands in the lanes containing RNA extracted from the *rpo*C2::*aad*A transformant and the MluI strain is approximately equal suggesting that these two strains have equal transcript levels of both *psb*C and *rrn*S. Slight variations in intensity may be due to unequal loading of RNA samples. This suggests that reduced levels of the *rpo*C2 gene have no effect on the transcription of these two chloroplast genes. The intensity of the bands in the lane containing WT RNA is higher, due to reduced transcription in cells grown in spc.

3.2.12 Effect of RNA polymerase inhibitors on growth of C. reinhardtii

Heteroplastic transformants are difficult to analyse in terms of phenotypic effects as they may maintain sufficient WT gene copies to have no detrimental effects on the cell. An alternative strategy to gene manipulation is to study the effects of RNA polymerase inhibitors on cells. A range of chloroplast RNA polymerase inhibitors have been used in higher plants and the effects of these on growth of whole *C. reinhardtii* cells was tested. Rifampicin, streptolydigin and tagetitoxin are inhibitors of PEP and *E. coli* RNA polymerase. Actinomycin D is a general RNA polymerase inhibitor and can be used as a control. WT *C. reinhardtii* CC-1021 cells were grown in TAP and HSM supplemented with i) actinomycin D at 40μg/ml ii) rifampicin at 250μg/ml iii) streptolydigin at 100μg/ml iv) tagetitoxin at 0.1μM and v) methanol at 2.5% ("no inhibitor" control) and the cells count determined every 24 hours until the controls reached stationary phase (~1.5 x 10^7 / ml). Fig. 3.10 (a) and 3.10 (b) show the resulting growth curves. Actinomycin D and rifampicin have an inhibitory effect on growth of WT *C. reinhardtii* in both TAP and HSM, whereas tagetitoxin and streptolydigin at the concentrations tested have no effect on cell growth in either TAP or HSM.



Total RNA was prepared from (A) a WT strain grown in TAP, (B) the *rpo*C2::aadA transformant F and (C) the MluI strain grown in TAP+spc^{50}. The filter was probed with (i) the 0.6kbp fragment containing the *psb*C gene and (ii) the 4.8kbp fragment containing the *rrn*S gene

Fig. 3.9 Northern analysis of *rpo*C2::*aad*A transformants

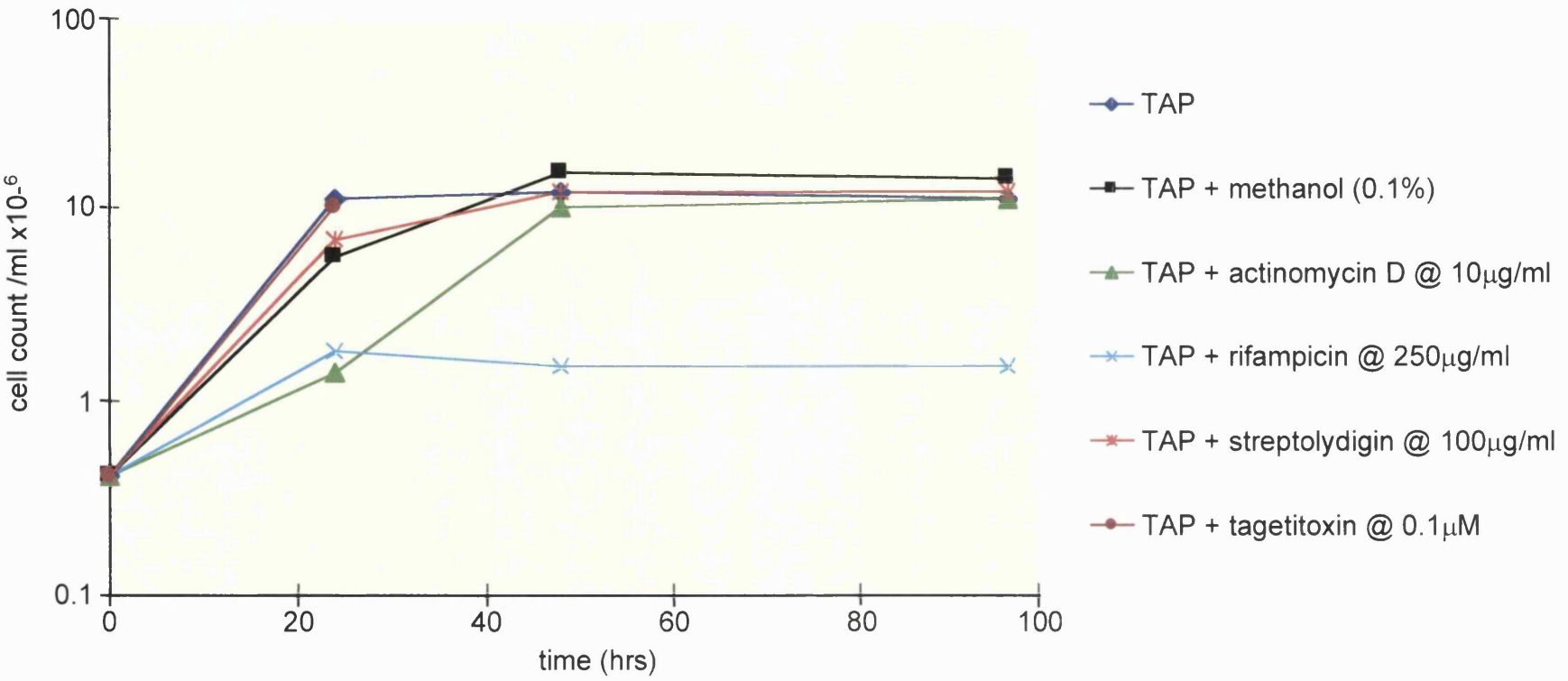


Fig. 3.10 (a) Effect of RNA polymerase inhibitors on growth of *C. reinhardtii* WT CC-1021 in heterotrophic growth conditions

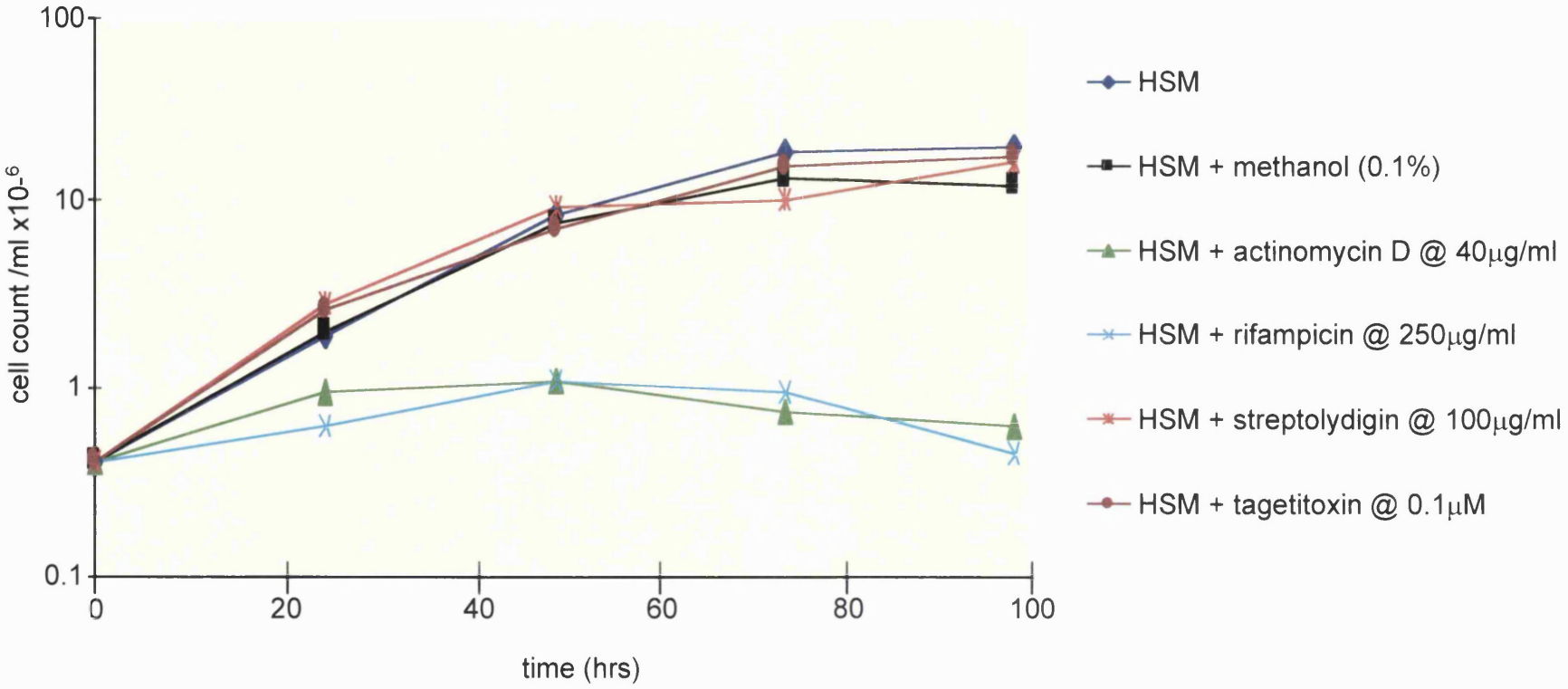


Fig. 3.10 (b) Effect of RNA polymerase inhibitors on growth of *C. reinhardtii* WT CC-1021 in phototrophic growth conditions

In order to investigate whether tagetitoxin and streptolydigin have an effect on cell wall deficient *C. reinhardtii* the experiment was repeated using *C. reinhardtii cw*10 CC-849 cells. This strain has a nuclear mutation, *cw*10, that results in a deficiency in the proteinaceous cell wall (Harris, 1989). *C. reinhardtii cw*10 cells were grown in TAP and HSM supplemented with i) streptolydigin at 100μg/ml ii) tagetitoxin at 0.1μM. The growth curves are shown in fig. 3.11. Neither of these inhibitors have an effect on cell wall deficient *C. reinhardtii* at the concentrations tested.

The minimum inhibitory concentration of rifampicin on growth of *C. reinhardtii* WT CC-1021 cells in TAP and HSM was determined. *C. reinhardtii* WT CC-1021 cells were grown in TAP and HSM supplemented with rifampicin at i) 25μg/ml ii) 50μg/ml iii) 100μg/ml iv) 150μg/ml v) 200μg/ml vi) 250μg/ml. Fig. 3.12 (a) and 3.12 (b) shows the resulting growth curves. The minimum inhibitory concentration of rifampicin in both TAP and HSM is ~100μg/ml.

3.2.13 Effect of rifampicin on fluorescence emission spectra

In order to investigate the effect of rifampicin on PSI, PSII and the LHCs the fluorescence emission spectra between 500 and 750nm was measured. When *C. reinhardtii* cells are exposed to light of 435nm normally assembled and functional PSI, PSII and the LHCs fluoresce at defined wavelengths. PSI fluoresces at 720nm and PSII and the LHCs fluoresce at 680nm (Kruse *et al.*, 1999). The intensity of the peaks indicates the relative abundance of the complexes. As rifampicin is an intense red / orange colour cells growing in rifampicin in liquid culture will not be exposed to the full range of wavelengths of light of cells growing in the absence of rifampicin. The red / orange of the rifampicin will act as a filter, blocking blue light from reaching the cells. As light affects chloroplast function at a range of levels from transcription (Christopher *et al.*, 1992) to photosynthetic functions it was necessary to ensure that the effects on PSI, PSII and the LHCs were due to the inhibition of the RNA polymerase by rifampicin, rather than a result of a lack of blue

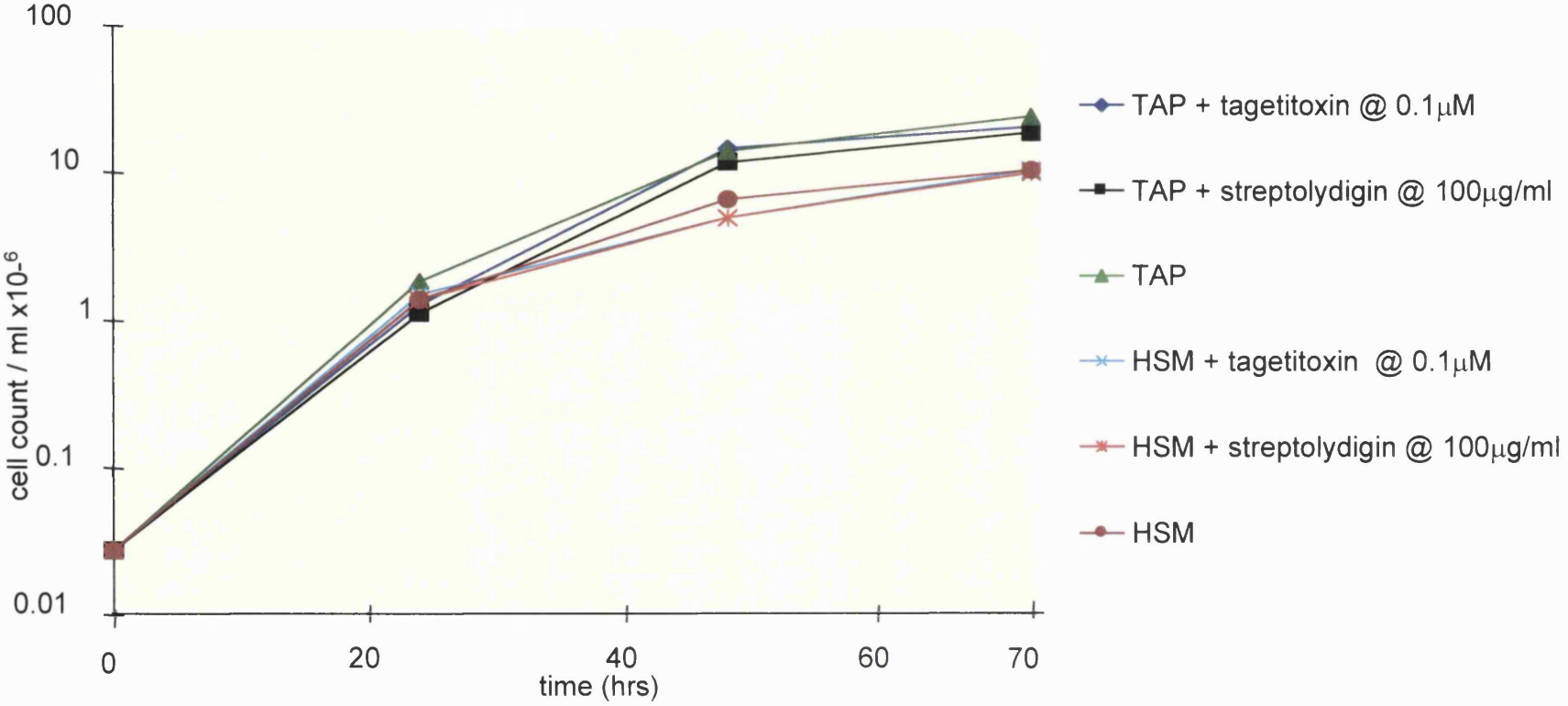


Fig. 3.11 Effect of RNA polymerase inhibitors on growth of cell wall deficient *C. reinhardtii cw*10 CC-849 in heterotrophic and phototrophic growth conditions

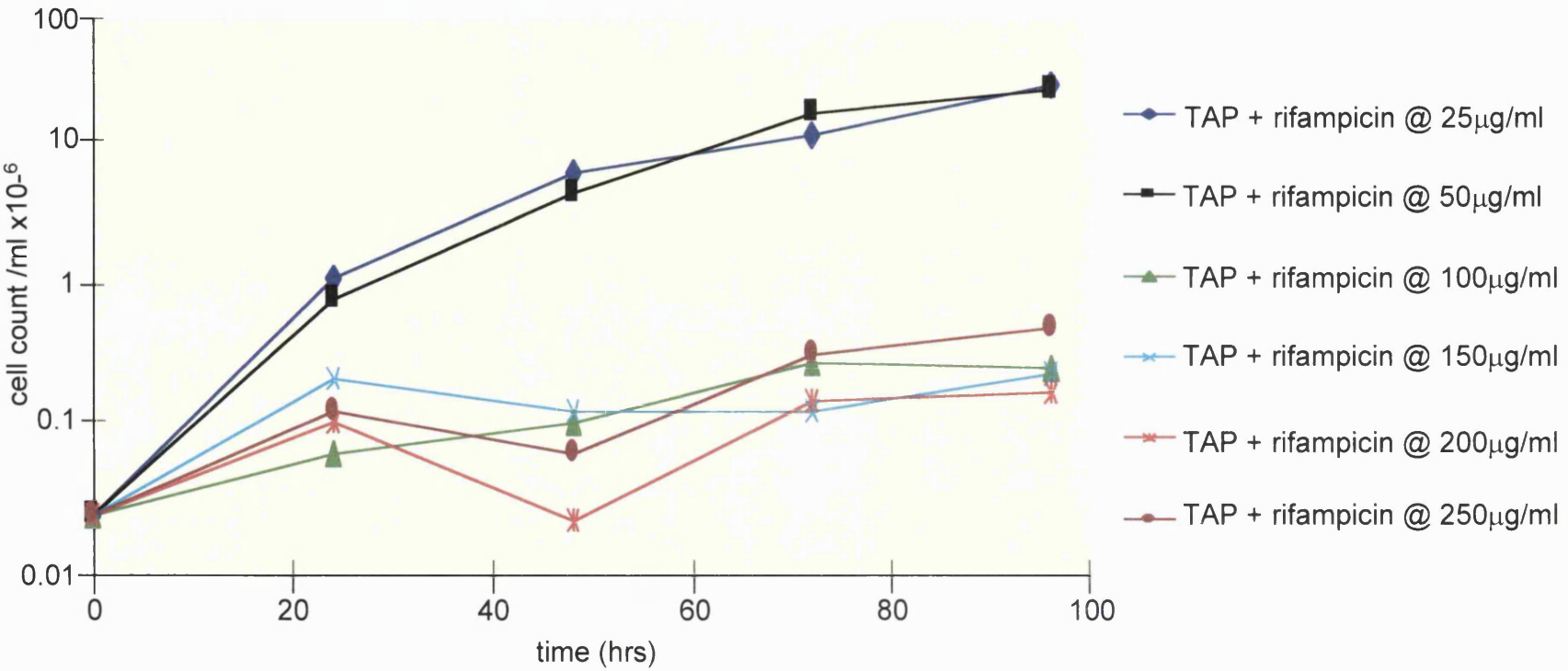


Fig. 3.12 (a) Minimum inhibitory concentration of rifampicin on growth of *C. reinhardtii* WT CC-1021 in heterotrophic growth conditions

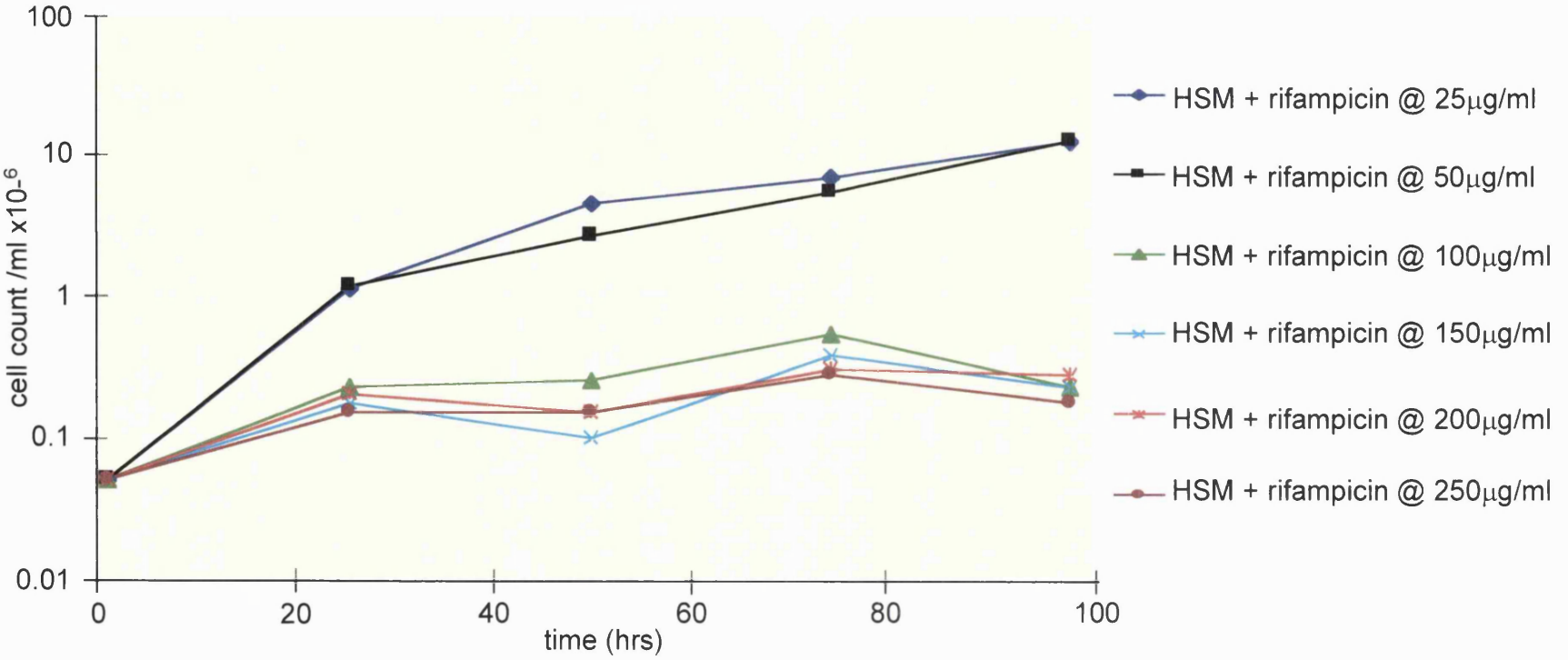


Fig. 3.12 (b) Minimum inhibitory concentration of rifampicin on growth of *C. reinhardtii* WT CC-1021 in phototrophic growth conditions

light on the cells. The absorbance spectrum of rifampicin at 100μg/ml was measured and compared to an orange filter. The absorbance spectra are shown in fig. 3.13. The filter absorbs light over the same wavelengths as rifampicin. *C. reinhardtii cw*10 CC-849 cells were grown to mid-log phase in HSM (~7.5 x 10^5 / ml) then rifampicin was added to a final concentration of 100μg/ml. *C. reinhardtii cw*10 cells were also grown in HSM under the orange filter and the fluorescence emission spectra between 500 and 750nm was measured and compared to untreated cells and cells incubated in rifampicin. The fluorescence emission spectra after 24 hours are shown in fig. 3.14 (a) and after 48 hours are shown in fig. 3.14 (b).

In the untreated control the peak at 680nm corresponds to normally assembled and functioning PSII and LHC. The peak at 720nm corresponds to normally assembled and functioning PSI. The intensity of the 680nm and 720nm peaks is approximately equal, suggesting that there are similar amounts of PSI and PSII, as expected. Untreated cells which are actively photosynthesising under white light will contain similar amounts of PSI and PSII. In the sample treated with rifampicin after 24 hours the peak at 680nm, corresponding to PSII and LHC, is less intense than the 720nm peak, corresponding to PSI. This suggests that there is less PSII in the rifampicin treated sample relative to PSI. In addition the 680nm peak is shifted to a lower wavelength, suggesting that there is less PSII relative to LHC. The 720nm peak is also shifted to a lower wavelength, this could be due to structural changes in PSI resulting from the inhibition of the chloroplast RNA polymerase by rifampicin. After the cells had been incubated in rifampicin for 48 hours the effects were even more marked. The intensity of the 680nm peak was much lower than the intensity of the 720nm peak, suggesting that the levels of PSII relative to PSI had dropped further. The 680nm and 720nm peaks were still shifted to the left. In contrast in the sample grown under the orange filter the 720nm peak is less intense than the 680nm peak, suggesting that there is less PSI than PSII. This suggests that the effects seen in the

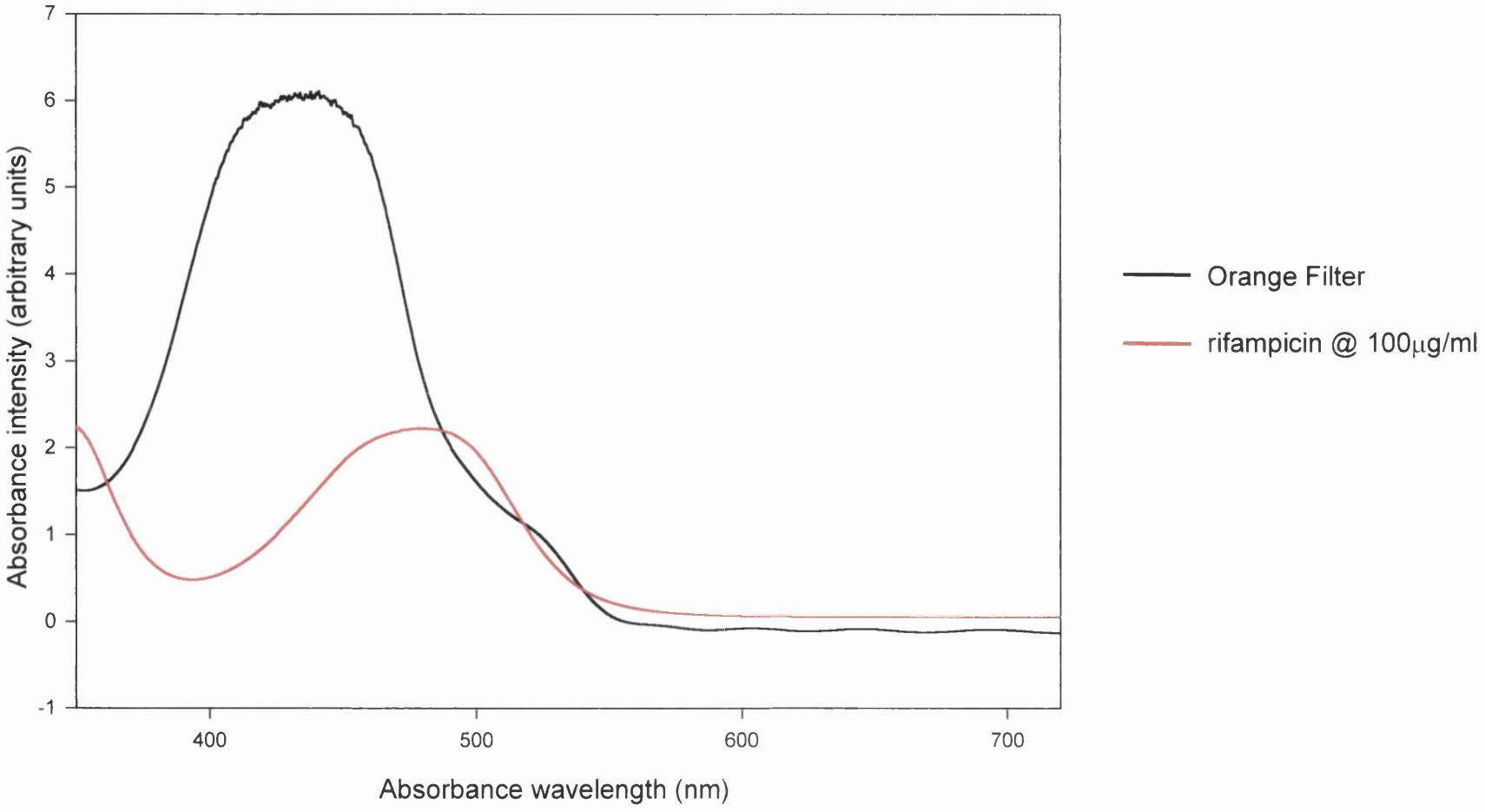


Fig. 3.13 Absorbance spectra of rifampicin @ 100μg/ml and an orange filter

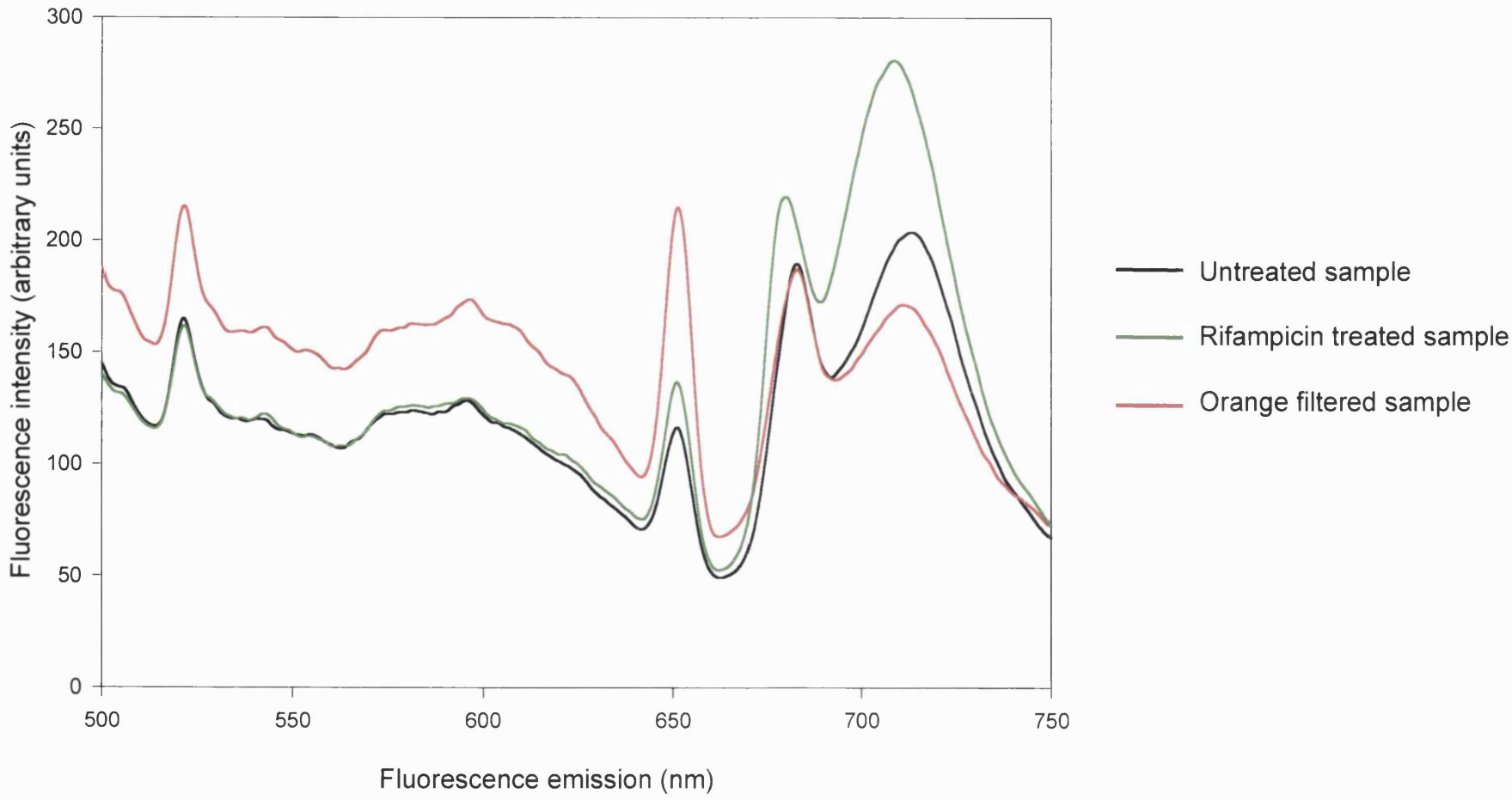


Fig. 3.14 (a) Fluorescence emission spectra of *C. reinhardtii cw*10 CC-849 cells after 24 hours at 77K

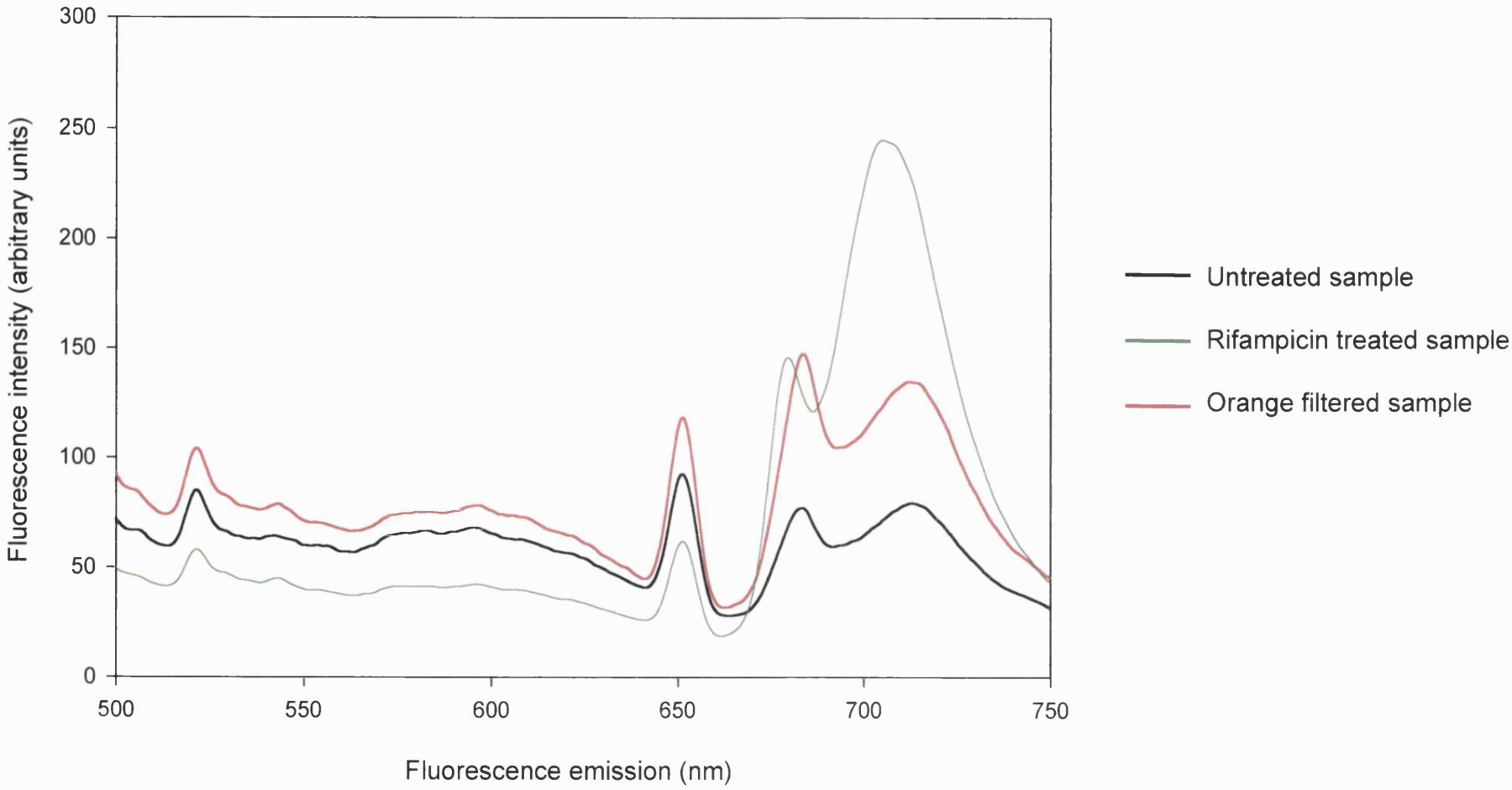


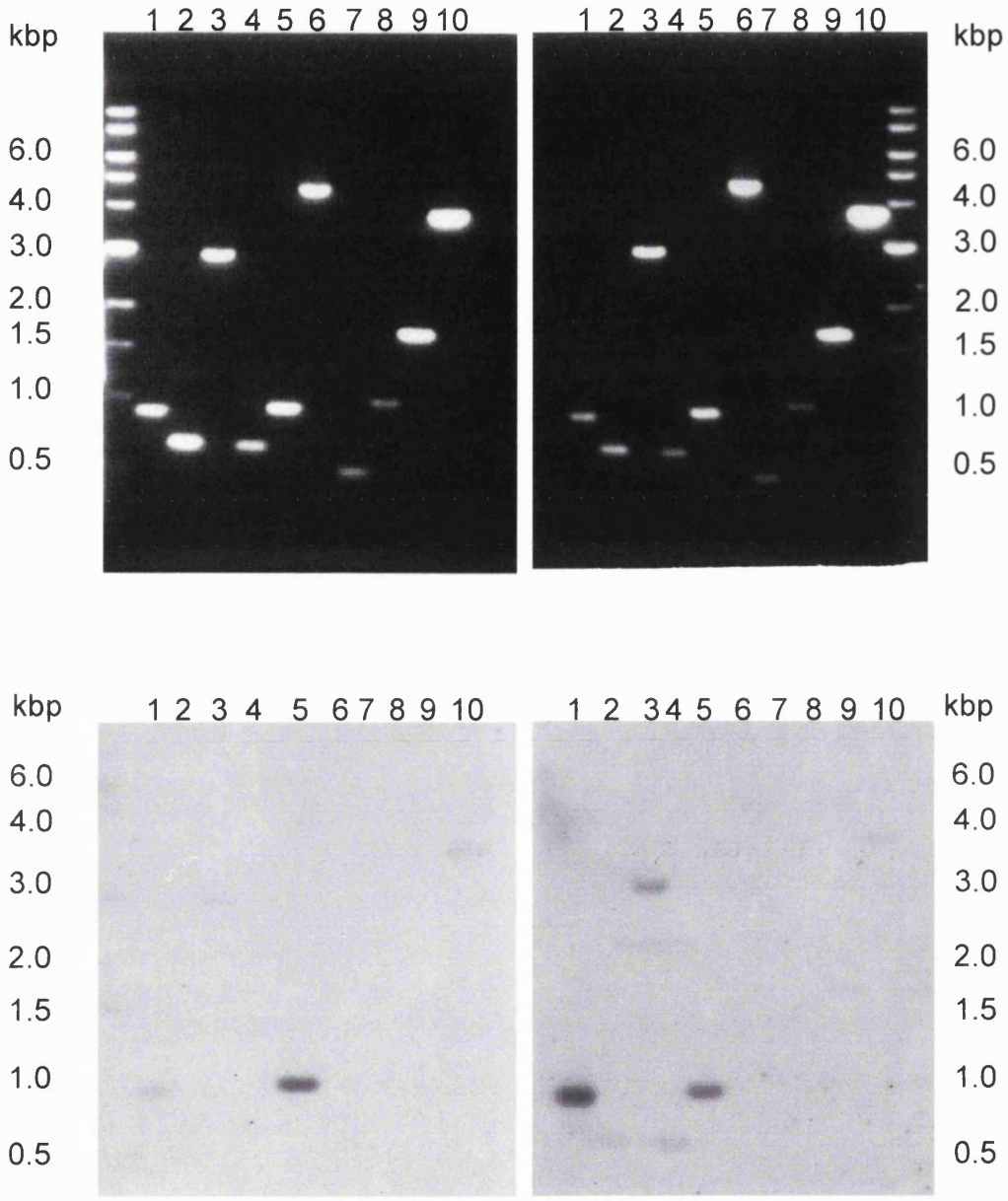
Fig. 3.14 (b) Fluorescence emission spectra of *C. reinhardtii cw*10 CC-849 cells after 48 hours at 77K

samples treated with rifampicin are due to an inhibition of the chloroplast RNA polymerase rather than a light effect.

3.2.14 Effect of rifampicin on transcription

In order to investigate the effect of rifampicin on transcription of a range of *C. reinhardtii* chloroplast, mitochondrial and nuclear genes an *in vivo* pulse labelling assay was carried out in the presence of rifampicin at 100μg/ml. The cells were pre-incubated in rifampicin prior to the pulse labelling for one hour. This is because rifampicin can only inhibit transcription initiation; elongation can occur in the presence of this drug. A one hour pre-incubation ensured that all transcripts were complete before the pulse labelling. The RNA was pulse radiolabelled with ^{32}P-UTP as described in 2.9.2, extracted and used to probe a Southern blot that had a variety of chloroplast, mitochondrial and nuclear gene fragments immobilised to it. Only transcripts synthesised in the presence of rifampicin will incorporate ^{32}P-UTP and be detectable on the autoradiogram. The resulting autoradiogram was compared to an untreated control. Both autoradiograms, and the agarose gels of the corresponding Southern blots are shown in fig. 3.15.

In the absence of rifampicin, transcripts can be detected for the following chloroplast genes; *rbc*L, encoding the large subunit of Rubisco, *atp*A, encoding a subunit of ATP synthase, *psa*A, encoding a subunit of PSI, *psb*C, encoding a subunit of PSII and *rrn*S, encoding 16S rRNA. Transcripts of a nuclear gene encoding rRNA can also be detected. Transcripts for the chloroplast genes *rpo*C2, encoding a subunit of PEP, *cem*A, encoding an envelope membrane protein, *orf*2971, encoding a putative nucleoid binding protein and the mitochondrial encoded gene *cox*1, encoding a subunit of cytochrome *c* oxidase can not be detected by this assay. In the presence of rifampicin only transcripts for the nuclear rRNA gene and for the chloroplast rRNA gene can be detected at levels comparable to the untreated control. Transcripts for the chloroplast *rbc*L and *psa*A genes are present at vastly reduced levels. This is summarised in Table 3.1.



Lane	Transcript	Lane	Transcript
1	*rbc*L	6	*rpo*C2
2	*atp*A	7	*cem*A
3	*psa*A	8	*orf*2971
4	*psb*C	9	*cox*1
5	*rrn*S	10	nrRNA

Fig. 3.15 *In vivo* transcription assay

Table 3.1 Results of the transcription assay

Gene	Product	Transcript detectable	Transcript detectable with rifampicin
***rbc*L (chloroplast)**	Subunit of Rubisco	✓	(✓)
***atp*A (chloroplast)**	Subunit of ATP synthase	✓	×
***psa*A (chloroplast)**	Subunit of PSI	✓	(✓)
***psb*C (chloroplast)**	Subunit of PSII	✓	×
***rrn*S (chloroplast)**	Ribosomal RNA	✓	✓
***rpo*C2 (chloroplast)**	Subunit of PEP	×	×
***cem*A (chloroplast)**	Envelope membrane protein	×	×
***orf*2971 (chloroplast)**	Putative nucleoid binding protein	×	×
***cox*1 (mitochondrial)**	Subunit of cytochrome *c* oxidase	×	×
nrRNA (nuclear)	Ribosomal RNA	✓	✓

It can be concluded that rifampicin at 100μg/ml does not inhibit nuclear transcription and that transcription of the *rrn*S gene appears to be insensitive to rifampicin. It is possible that there is a rifampicin insensitive RNA polymerase in *C. reinhardtii* chloroplasts.

3.3 Discussion

3.3.1 rpo*C2 reverse genetics*

The aim of this chapter was to characterise the PEP of *C. reinhardtii* by carrying out reverse genetics on one of the genes encoding an essential subunit. Re-sequencing of the *rpo*C2 gene, revealed that *rpo*C2 extends as a single large open reading frame through orf472 to give a predicted protein of 3120 amino acids containing large insertion sequences which do not show homology to the β″ subunit of PEP (Nuotio, S. & Purton, S. unpublished

data). It was reported that disruption of the *rpo*C2 gene identified by Fong and Surzycki and of orf472 resulted in heteroplastic transformants (Rochaix, 1997), (Goldschmidt-Clermont, 1991) but gene disruption experiments have not been carried out since the correct sequence was determined. *C. reinhardtii* mutants in which a portion was removed from the middle of the *rpo*C2 gene and replaced with the *aad*A cassette had been made previously in this laboratory, but they had not been analysed. Southern analysis confirmed that these mutants were heteroplastic. The disrupted form of *rpo*C2 had integrated into some copies of the chloroplast genome but the cell retains WT gene copies. This suggests that the *rpo*C2 gene, and hence PEP, is essential to cell viability in *C. reinhardtii.* This is in contrast to the findings of Allison and co-workers in tobacco where *rpo*B disruption mutants were found to be homoplastic (Allison *et al.*, 1996). The hypothesis that *rpo*C2 is essential was reinforced by the finding that in the absence of a selective pressure for the disrupted form of the gene (ie in the absence of spc) the cell rapidly replaces its disrupted gene copies with functional gene copies. This suggests that there is a very strong selective pressure for functional *rpo*C2. Spot tests of the *rpo*C2::*aad*A transformants confirmed that they were photosynthetic and resistant to spc. The disruption appears to have no detrimental effect on cell growth as compared to WT *C. reinhardtii.*

The work of Allison and co-workers, however, differed fundamentally from the gene disruptions carried out in this laboratory and others on *C. reinhardtii* PEP genes. The *aad*A cassette used by Allison and co-workers was driven from the promoter of the *rrn*S gene, a housekeeping gene encoding for 16S rRNA (Svab & Maliga, 1993), whereas the *aad*A cassette used in the *C. reinhardtii* work was driven from the promoter of the *atp*A gene, a photosynthetic gene encoding for a subunit of ATP synthase (Goldschmidt-Clermont, 1991). Given that in higher plants PEP transcribes photosynthetic genes and NEP transcribes housekeeping genes the heteroplastic state observed in the *C. reinhardtii* gene knockouts could be because PEP is needed to transcribe the *aad*A cassette, but not for cell viability *per se*. In

order to clarify this the *atp*A promoter of the *C. reinhardtii aad*A cassette was replaced with the promoter of the *C. reinhardtii rrn*S gene. This new *aadA* cassette (*rrn*S*aad*A) was used to disrupt *rpo*C2 as before, and the construct was used to transform WT *C. reinhardtii* using the biolistic method. Although the *rrn*S*aad*A cassette confers high levels of resistance to spc in *E. coli* (up to 500μg/ml), *C. reinhardtii rpo*C2 disruption transformants could not be isolated. Spc^R transformants were obtained, but Southern analysis revealed that only WT copies of *rpo*C2 were present and spot tests confirmed that the resistance was due to a spontaneous mutation rather than the integration of the *aad*A cassette as the transformants were resistant to spc but not str.

The upstream region of the *rrn*S gene which was amplified by PCR provide the new promoter for the *aad*A cassette contains –10 and –35 sequences which resemble prokaryotic promoters and a transcription start site. Although the transcription start site has been mapped by S1 nuclease digestion the possibility remains that the transcript has been processed and the true transcription start site is further upstream. Also, the region upstream of the tobacco *rrn*S gene, and the *rrn*S genes of other plants are well conserved. The upstream region of the *C. reinhardtii rrn*S gene differs from that of higher plants in that it does not contain a $tRNA^{Val}$ gene which is located 300bp upstream of the *rrn*S gene in higher plants (Schneider *et al.*, 1985). Also the new *rrn*S promoter was a chimeric structure. As *rrn*S encodes for ribosomal RNA rather than protein it may lack elements upstream of the start site needed for efficient translation. For these reasons the promoter of the new *rrn*S*aad*A cassette could be non-functional in *C. reinhardtii*. This could be tested by using this cassette to knock out a gene which can readily be disrupted such as *psb*H (O'Connor *et al.*, 1998) or inserting the cassette in a non-coding region (such as downstream of *psb*H), then transforming a Δ*psb*H mutant (Bateman & Purton, 2000). Transformants can be selected for under phototrophic conditions and spot tests on increasing concentrations of spc can be used to determine whether the marker is expressed (fig. 3.16).

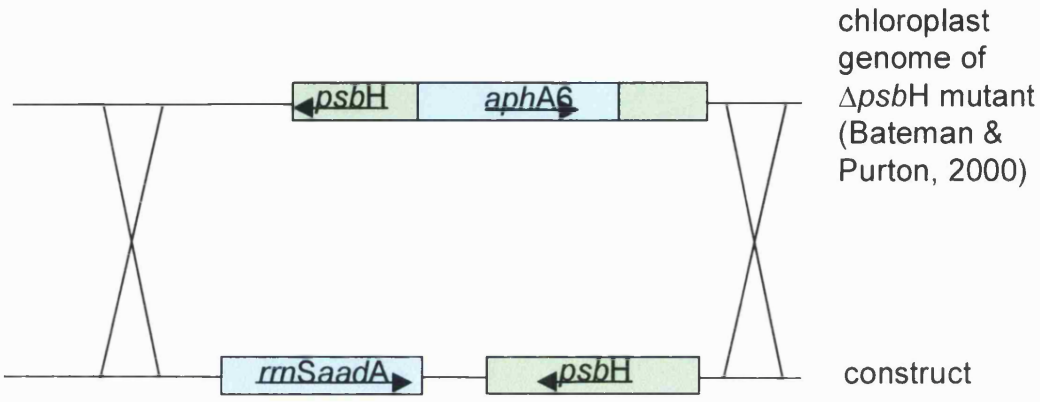


Fig. 3.16 An experiment to test the expression of the *rrnSaadA* cassette. Transformation of an Δ*psb*H mutant with a construct containing the *rrnSaadA* cassette inserted in a non-coding region allows the transformants to be selected independently of the cassette. The transformants can subsequently be analysed to determine whether the cassette is transcribed

As an alternative strategy a co-transformation approach was adopted. In *C. reinhardtii* non-essential genes, such as the photosynthetic genes *atp*B (Kindle *et al.*, 1991) and *rbc*L (Newman *et al.*, 1991) can be disrupted using this method. Homoplastic disruption mutants can be obtained by bombarding *C. reinhardtii* cells with two plasmids, one containing the gene of interest disrupted with a "stuffer", non-expressed piece of DNA, and a second containing a mutated form of the *rrn*S gene which confers resistance to spectinomycin. Disruption of potentially essential genes using this method has not been reported, but three possible outcomes were predicted: i) *rpo*C2 is not essential and the heteroplastic state observed previously is an artefact of disruption using an expressed cassette driven from a photosynthetic gene promoter so homoplastic disruption transformants could be isolated by a co-transformation method; ii) *rpo*C2 is essential and heteroplastic transformants are isolated by a co-transformation method; iii) *rpo*C2 is essential, but the selective pressure for functional *rpo*C2 is so strong that the cell will not tolerate a disruption in the absence of a selective pressure for that disruption. A large number of transformants were screened and, even at the first round of growth, no disrupted copies of *rpo*C2 could be detected. This suggests that disrupting an essential gene by a co-transformation has the third result. It was hoped that disruption of *rpo*C2 with a non-expressed piece of stuffer DNA would overcome the dependence on the expression of the *aad*A cassette by PEP, however, it seems that the selective pressure for functional *rpo*C2 is so strong that it will not tolerate this disruption. This is consistent with the previous findings that *rpo*C2::*aad*A heteroplastic transformants grown in the absence of spc rapidly replace their disrupted gene copies with functional gene copies.

The finding that *rpo*C2, and hence PEP, is essential to *C. reinhardtii* cell viability is interesting in that it represents a significant difference in the

transcriptional apparatus of higher plant and algal species. It has been proposed that the chloroplast NEP is a result of a gene duplication event from the nuclear encoded mitochondrial RNA polymerase as the corresponding proteins are both single subunit phage type polymerases and are strikingly similar in terms of amino-acid sequence (Gray & Lang, 1998). The duplication was followed by the acquisition of a chloroplast targeting sequence, allowing the protein to be re-routed to the chloroplast. Chloroplast genes then acquired the NEP promoters and an additional level of control of transcription, as well as control over the chloroplast by the nucleus was complete. It is unclear at which point in evolutionary time this duplication event occurred. Given the apparently essential nature of PEP in *C. reinhardtii* it seems likely that NEP was acquired after the evolution of higher plants. Searches of the *C. reinhardtii* EST database have not identified a potential NEP gene (4.2.13).

In order to investigate this further, *rpo* gene knockouts need to be made in other algal species. Recently, advances have been made in the genetic transformation of several classes of algae. The chloroplast genome of the green alga *Euglena gracilis* has been successfully transformed (R. Hallick, pers. comm.). The entire chloroplast genome of this organism has been sequenced and the *rpo* genes identified (Hallick *et al.*, 1993). The *aad*A cassette has recently been developed for use in this alga (Hallick, pers. comm.). The chloroplast genome of the red alga *Porphyridium* has also been transformed using microparticle bombardment and a selectable marker based on resistance to the herbicide sulfomethuron methyl (SMM) is being developed. A resistant variant of the gene encoding the Acetohydroxyacid synthase (AHAS) enzyme, which is the target enzyme of SMM has been cloned and used to transform wild type *Porphyridium* (Lapidot *et al.*, 1999). The nuclear genomes of the green alga *Volvox carteri* and several diatom species including *Cyclotella cryptica*, *Navicula saprophila* and *Phaeodactylum tricornutum* have been successfully transformed by microparticle bombardment and of *Acetabularia acetabulum*, a giant unicellular green alga

by microinjection. Stable nuclear transformation methods have also been developed for *Dunaliella salina*, and transient expression of heterologous genes has been reported in the green algae *Chlorella ellipsoida*, *Chara corallina* and *Ulva lactuca* and the red algae *Kappaphycus alvarezii* and *Porphyra miniata* (Stevens & Purton, 1997). As microparticle bombardment has been used successfully in a number of algal species, it could be used as a general procedure for delivery of DNA into the chloroplast of these species. Existing selectable cassettes such as *aadA* (Goldschmidt-Clermont, 1991) and *aphA6* (Bateman & Purton, 2000) could be used in other algal species provided that the *C. reinhardtii* regulatory sequences were replaced with those from the algal genome in question.

3.3.2 Effects of RNA polymerase inhibitors on C. reinhardtii

Initially the effects of the RNA polymerase inhibitors actinomycin D, rifampicin, streptolydigin and tagetitoxin on growth of *C. reinhardtii* cells in liquid culture was tested. The effects on growth in both heterotrophic (TAP in the light) and phototrophic (HSM in the light) conditions were studied. If PEP activity is dispensable for cell viability, but essential for phototrophic growth, as in higher plants, then the cells should grow in heterotrophic conditions, but not in phototrophic conditions. Streptolydigin at 100μg/ml and tagetitoxin at 0.1μM have no effect on WT or cell wall deficient *C. reinhardtii*. As these inhibitors specifically inhibit PEP, the most likely explanations for this is that the concentrations used in these experiments are too low to have an inhibitory effect *in vivo* or that the inhibitors can not penetrate the cell or the chloroplast. Rifampicin, however, inhibits cell growth in both heterotrophic and phototrophic conditions at 100μg/ml. This also suggests that PEP is essential for cell viability as NEP is not sensitive to rifampicin.

The fluorescence emission spectra allows the effect of rifampicin on PSI, PSII and the LHCs to be studied. The amount of PSII relative to PSI in the rifampicin treated sample is lower than in the untreated sample, as measured

by the intensities of the 680nm and 720nm peaks. PSII turns over faster than PSI, so blocking transcription would have an immediate effect on levels of PSII, whereas the levels of PSI would drop more slowly. This also explains the drop in levels of PSII relative to LHC, as measured by the shift of the 680nm peak to a lower wavelength. Within PSII the D1 polypeptide, encoded by the *psb*A gene, is turned over most rapidly, so it seems likely that rifampicin has an effect on the transcription of *psb*A. Although PSI turns over more slowly than PSII (Aro *et al.*, 1993) rifampicin has still had an effect as the 720nm peak is also shifted to a lower wavelength. This shift is characteristic of structural changes in PSI, and could be the result of rifampicin blocking transcription of a chloroplast encoded structural subunit of PSI which is turned over more quickly than the other subunits. A possible candidate for this is the protein encoded by *psa*C which has two iron-sulphur (Fe-S) centres and is, therefore, very sensitive to O_2, which can oxidise the Fe-S centres. The PsaC protein is probably turned over more rapidly than the other proteins in PSI due to damage to the Fe-S centres. These effects are different to those observed in the sample grown under an orange filter where there is more PSII relative to PSI as measured by the relative intensities of the 680nm and 720nm peaks. PSII uses the shorter wavelengths of light that are blocked out by the orange filter so more PSII is produced to compensate for this. The effects observed in the rifampicin treated sample are a result of the inhibition of the RNA polymerase, rather than an effect of the light.

Finally the effect of rifampicin on transcription was studied. A transcription assay was developed in which RNA transcripts are pulse labelled with ^{32}P-UTP *in vivo*. The RNA transcripts are then extracted from the cells and used to probe a Southern blot which has a variety of chloroplast, mitochondrial and nuclear gene fragments immobilised to it. The transcripts will hybridise to the gene fragments and those which have been transcribed during the pulse labelling can be visualised by autoradiography. This was carried out in the presence of rifampicin in order to determine which genes are transcribed in the presence of this drug. As expected the transcription of nuclear genes is

unaffected by rifampicin as a transcript can be detected for the nuclear rRNA gene. The effect of rifampicin on transcription of the mitochondrial genome could not be determined as this assay is not sensitive enough to detect the *cox*1 transcript, encoding a subunit of cytochrome *c* oxidase, in the presence or absence of rifampicin. The assay was not sensitive enough to detect the chloroplast *rpo*C2 (encoding a subunit of PEP) *cem*A (encoding an envelope membrane protein) or *orf*2971 (encoding a putative nucleoid binding protein) transcripts. This is unsurprising as neither *rpo*C2 nor *orf*2971 can be detected by northern analysis and are poorly expressed genes. The chloroplast *rbc*L (encoding a subunit of Rubisco), *atp*A (encoding a subunit of ATP synthase), *psa*A (encoding a subunit of PSI), *psb*C (encoding a subunit of PSII) and *rrn*S (encoding a rRNA) transcripts could be detected. Only the transcript of the *rrn*S gene could be detected in the presence of rifampicin at levels comparable to the untreated control.

The effect of rifampicin on transcriptionally active extracts from higher plant and algal chloroplasts has been tested and given conflicting results. Bottomley and co-workers found that rifampicin had no effect on chloroplast transcription in pea, maize, radish and spinach neither *in vivo* nor *in vitro* (Bottomley *et al.*, 1971) whereas Surzycki found that the chloroplast RNA polymerase of *C. reinhardtii* was inhibited *in vivo* and *in vitro* by this drug (Surzycki, 1969). Despite reports that the chloroplast RNA polymerase of *C. reinhardtii* is completely inhibited by rifampicin as measured by incorporation of NTPs into chloroplast RNA (Guertin & Bellemare, 1979) Jahn found that the *C. reinhardtii* chloroplast $tRNA^{Glu}$ gene was transcribed in the presence of rifampicin (Jahn, 1992). In mustard distinct rifampicin sensitive and insensitive RNA polymerase activities were identified in etioplasts and chloroplasts. In etioplasts the rifampicin sensitive B enzyme predominates, whereas in chloroplasts the rifampicin insensitive A enzyme is the principle activity (Pfannschmidt & Link, 1994). Both A and B enzymes transcribe all gene classes *in vitro* and rifampicin inhibits transcription of all gene classes by the B enzyme. However, transcript levels of *psb*A and *rbc*L *in vivo* are reduced in seedlings grown in rifampicin whereas transcript levels of *rps*16,

*trn*G, *rrn* and *rpo*B are unaffected (Pfannschmidt & Link, 1997). It has recently been shown that the A and B enzymes are, in fact, structurally related and can be interconverted by phosphorylation of the B enzyme or dephosphorylation of the A enzyme (Pfannschmidt *et al.*, 2000).

It was hoped that these inhibitor studies would provide evidence for a second RNA polymerase activity in *C. reinhardtii* chloroplasts. The clear demonstration of a rifampicin insensitive RNA polymerase activity in *C. reinhardtii* chloroplasts which specifically transcribes housekeeping genes would indicate the presence of a NEP activity. However, the inhibitory effect of rifampicin on PEP is poorly understood and it appears that PEP exists in rifampicin sensitive and insensitive forms. It is possible that the *rrn*S gene is being transcribed by a non-phosphorylated form of PEP. To overcome this a more specific PEP inhibitor, such as tagetitoxin, could be used in an *in vivo* transcription assay if an inhibitory effect can be seen on cell growth. A higher concentration of tagetitoxin, for example, may have an inhibitory effect on growth of *C. reinhardtii.*

Chapter 4 – Cloning of *C. reinhardtii* Nuclear Genes Involved in Chloroplast Transcription

4.1 Introduction

Recent research has identified nuclear genes from higher plants and red algae whose products are targeted to the chloroplast and function in transcription of the chloroplast genome. Genes resembling prokaryotic RNA polymerase σ factors have been cloned and sequenced from the nuclear genomes of the red alga *C. caldarium* (Tanaka *et al.*, 1996), (Oikawa *et al.*, 1998) (Liu *et al.*, 1999), (Liu & Troxler, 1996) and the plants *Arabidopsis* (Isono *et al.*, 1997), (Tanaka *et al.*, 1997), rice (Tozawa *et al.*, 1998), mustard (Kestermann *et al.*, 1998), maize (Tan & Troxler, 1999) and wheat (Morikawa *et al.*, 1999) as discussed in 1.4.6. In prokaryotes the σ subunit binds to the RNA polymerase core enzyme and promotes accurate promoter recognition by the holoenzyme. In plant and algal species the genes encoding these σ factors are transcribed in response to light. Genes encoding σ factors have not been identified in green algal species, although preliminary biochemical analysis using *C. reinhardtii* has identified two proteins with σ like activity (Surzycki & Shellenbarger, 1976).

Genes showing strong amino acid sequence similarity to mitochondrial and bacteriophage T7 RNA polymerases have also been cloned from the nuclear genomes of *Arabidopsis* (Hedtke *et al.*, 1997), maize (Chang *et al.*, 1999) and wheat (Ikeda & Gray, 1999). These genes have chloroplast targeting sequences and in some cases the products have been shown to be chloroplast located as discussed in 1.4.7. It is proposed that these genes encode the NEP (Nuclear Encoded Polymerase) identified by Allison and co-workers following deletion of subunits of PEP in tobacco (Allison *et al.*, 1996). It has also been suggested that these NEP genes resulted from a gene duplication event from the nuclear encoded mitochondrial RNA polymerase and that this occurred at some point during the evolution of higher plants from within the green algal lineage (Gray & Lang, 1998). If so the question remains as to whether algal species also have a NEP.

The core subunits of the PEP are encoded by four plastid gene, *rpo*A, *rpo*B, *rpo*C1 and *rpo*C2. In *C. reinhardtii* the *rpo*A gene, encoding the α subunit of PEP, is absent from the chloroplast genome. As discussed in 1.3.4 and shown in Table 1.3 the *rpo*A gene is present in all plastid genomes sequenced to date except those of the green alga *Euglena gracilis* (Hallick *et al.*, 1993), and the protists *Plasmodium falciparum* (Wilson *et al.*, 1996) and *Toxoplasma gondii*. Given that the α subunit is an essential subunit of PEP it seems likely that the *rpo*A gene has been transferred to the nucleus in these species, and in the case of *Plasmodium falciparum*, a nuclear gene for *rpo*A has recently been identified (Sato *et al.*, 2000).

The aims of this chapter are;

- To clone and sequence genes encoding PEP σ factors from the nuclear genome of *C. reinhardtii*.
- To attempt to clone a gene encoding a NEP from the nuclear genome of *C. reinhardtii*.
- To attempt to identify and clone the *rpo*A gene from the nuclear genome of *C. reinhardtii*.

4.2 Results

4.2.1 Attempt to clone PEP σ factor genes by RACE

RACE (Rapid Amplification of cDNA Ends) was used in an attempt to clone genes for PEP σ factors from the nuclear genome of *C. reinhardtii*. Fig. 4.1 shows a general scheme of the RACE protocol. Degenerate primers were designed to conserved regions of known plant, algal and bacterial σ factors (fig. 4.2). PolyA mRNA was extracted from WT *C. reinhardtii* CC-1021 cells

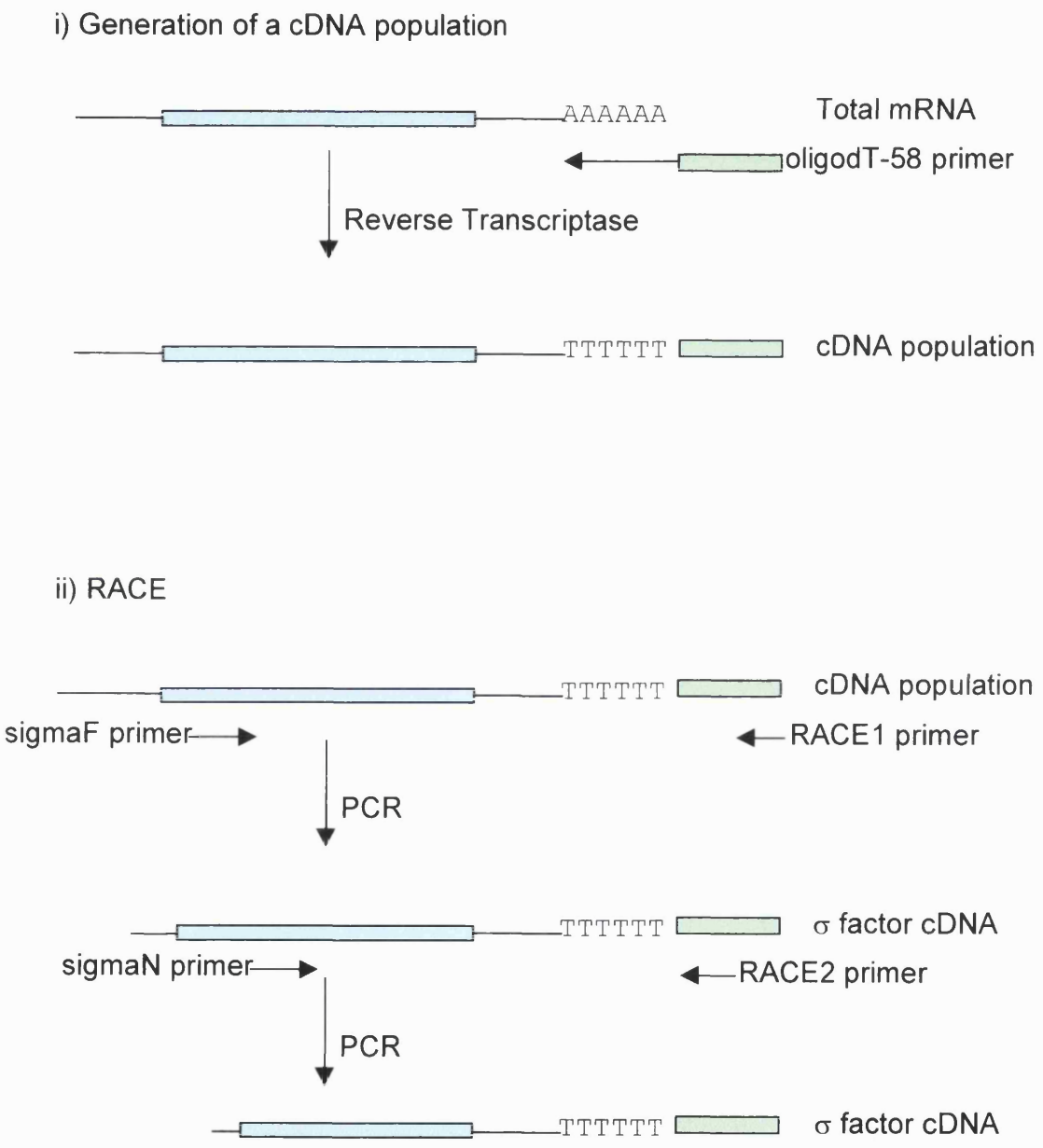


Fig. 4.1 Generation of a cDNA population, followed by RACE

Fig. 4.2 Alignment of portions of the protein sequences of σ factors from a range of species to show the location of the degenerate primers used in RACE

grown in TAP and converted to cDNA using the oligodT-58 primer and reverse transcriptase. A PCR amplification was carried out using the cDNA template and the RACE1 and Sigma(F) primers. The RACE product was then used as a template in a second amplification using the RACE2 primer, which is nested to the RACE1 primer, and the Sigma(N) primer, which is nested to the Sigma(F) primer. Despite several attempts to optimise PCR conditions, no products were recovered. As discussed at the end of this chapter this can be attributed to errors in the primer sequence and a lack of expression of σ factor genes under the conditions used.

4.2.2 Identification of PEP σ factor genes in the EST database

The *C. reinhardtii* nuclear genome is being characterised by an EST (Expressed Sequence Tags) approach (Asamizu *et al.*, 1999). ESTs represent partial sequences of cDNAs of randomly chosen clones. The EST for an interesting gene can then be used to screen a genomic library. In order to identify an EST corresponding to a PEP σ factor gene, a BLAST search of the EST database (http://www.ncbi.nih.gov:80/BLAST/) was carried out using the *E. coli rpo*D gene. Two *C. reinhardtii* ESTs with significant homology to the *E. coli* σ factor were identified (Accession numbers AV390342, AV392956). The Kazusa DNA Research Institute supplied these ESTs as plasmid clones. Fig. 4.3 shows the result of a BLAST search using the EST sequences. The AV390342 clone is most similar to a σ factor from the cyanobacterium *Synechococcus* and the AV392956 clone is most similar to σ factor from *Arabidopsis*.

Table 4.1 EST clone numbers and accession numbers

Accession Number	Clone Number
AV390342	CM093d01
AV392956	CM058g10

AV390342 BLAST search result

```
C. reinhardtii: 73   PSEDDLLAKAAGLPVTRVQMLMETARSAASLDTPIGGNELGPTVKDSVEDEREAADEEFG 252
                     P+E+++  +   + + +++ + ++A+   SL+TPIG  E    + D +E + E  ++E
Synechococcus : 251  PTEEEIATRME-MTIEKLRFIAKSAQLPISLETPIGKEE-DSRLGDFIEADGETPEDEVA 308

C. reinhardtii: 253  SDSLRNDMEAMLLELPEREARVVRLRFGLDDGKEWTLEEIGEALNVTRERIRQIEAKALR 432
                      + LR D+E +L  L  RE  V+RLR+GLDDG+  TLEEIG+  NVTRERIRQIEAKALR
Synechococcus : 309  KNLLREDLEGVLSTLSPRERDVLRLRYGLDDGRMKTLEEIGQLFNVTRERIRQIEAKALR 368

C. reinhardtii: 433  KLR 441
                     KLR
Synechococcus : 369  KLR 371
```

AV392956 BLAST search result

```
C. reinhardtii:  17  RDFLFLEKVKRQCEKTLHRPATSEEIAAAVAMDVESLKLRYDAGLKAKELLLKSNYKLVM 196
                     +D L LE+++ +  +   R  T  + A+A  +D +SL+ R   G   K+ ++KSN +LV+
Arabidopsis:     288 QDLLKLERLQTELTERSGRQPTFAQWASAAGVDQKSLRQRIHHGTLCKDKMIKSNIRLVI 347

C. reinhardtii: 197  TVCKSFVGKGPHIQDLVSEGVKGLLKGVEKYDATKGFRFGTYAHWWIRQAVSRSLAETGR 376
                     ++ K++ G G ++QDLV EG +GL++G EK+DATKGF+F TYAHWWI+QAV +SL++  R
Arabidopsis:     348 SIAKNYQGAGMNLQDLVQEGCRGLVRGAEKFDATKGFKFSTYAHWWIKQAVRKSLSDQSR 407

C. reinhardtii: 377  AVRLPMHMIEQLTRLKNLSAKLQTQLAR 460
                      +RLP HM+E   R+K    +L ++  +
Arabidopsis:     408 MIRLPFHMVEATYRVKEARKQLYSETGK 435
```

Fig. 4.3 BLAST search results from the EST clones. Amino acids in red are identical between the two species, amino acids in green represent conservative amino acid changes

4.2.3 Sequencing the EST clones

In order to confirm that the clones correspond to the *rpo*D gene and to complete the cDNA sequence the 1.4kbp insert from the AV390342 clone and the 3kbp insert from the AV392956 clone were sequenced with T3 and T7 primers, together with internal primers (as detailed in Appendix I). The sequencing revealed that the AV392956 clone in fact contains two cDNA inserts. One is a longer version of the *rpo*D cDNA of AV390342, the other corresponds to a gene of unknown function. BLAST searches using the sequence of this second gene failed to identify any similar genes. The complete sequence of the AV390342 clone was not determined as efforts were made to sequence the longer clone. Fig. 4.4 shows the physical maps of the two EST clones.

The 1.4kbp *Xho*I fragment from the AV392956 clone corresponding to the gene of unknown function was removed by RE digestion and the vector re-ligated to give the plasmid pSKrpoD25′. This plasmid was sequenced using the T3 and T7 primers, together with internal primers to produce a complete sequence of the insert corresponding to a partial cDNA sequence of *rpo*D (fig. 4.5).

4.2.4 Southern analysis of WT C. reinhardtii *genomic DNA using the EST clones*

In order to determine whether the genes corresponding to the EST clones are present as single copies, and to try to identify a gene family of σ factors in the nuclear genome of *C. reinhardtii* a genomic Southern blot was carried out. Total DNA was extracted from WT *C. reinhardtii* CC-1021 and digested with i) *Sma*I ii) *Pvu*II iii) *Hind*III iv) *Sph*I v) *Pst*I vi) *Hpa*I. The digested DNA was run on a 1% agarose gel and transferred to nylon by the Southern blotting technique. A radiolabelled probe was made from the 1.4kbp *Eco*RI-*Kpn*I fragment from AV390342 and used to probe the blot. Fig. 4.6 shows the

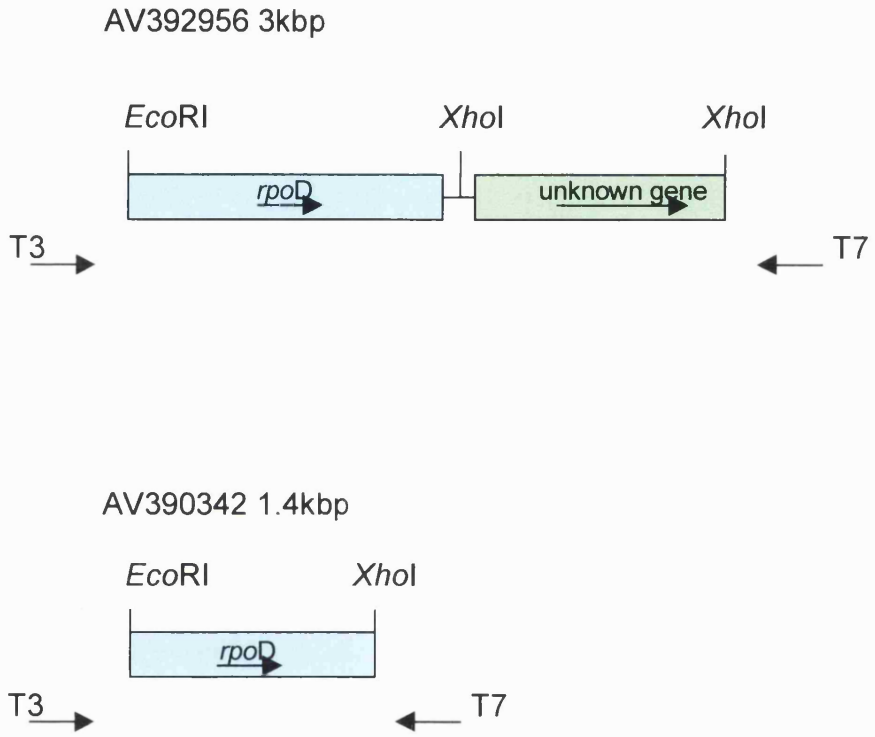
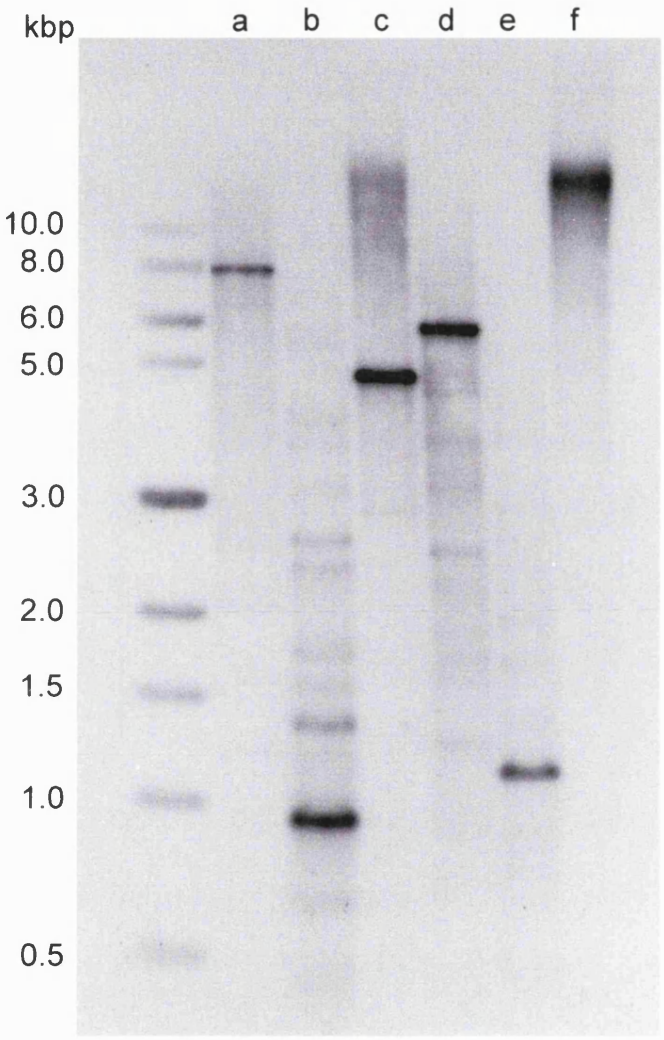


Fig. 4.4 Physical maps of the EST clones

Fig. 4.5 Sequence of the AV392956 cDNA insert



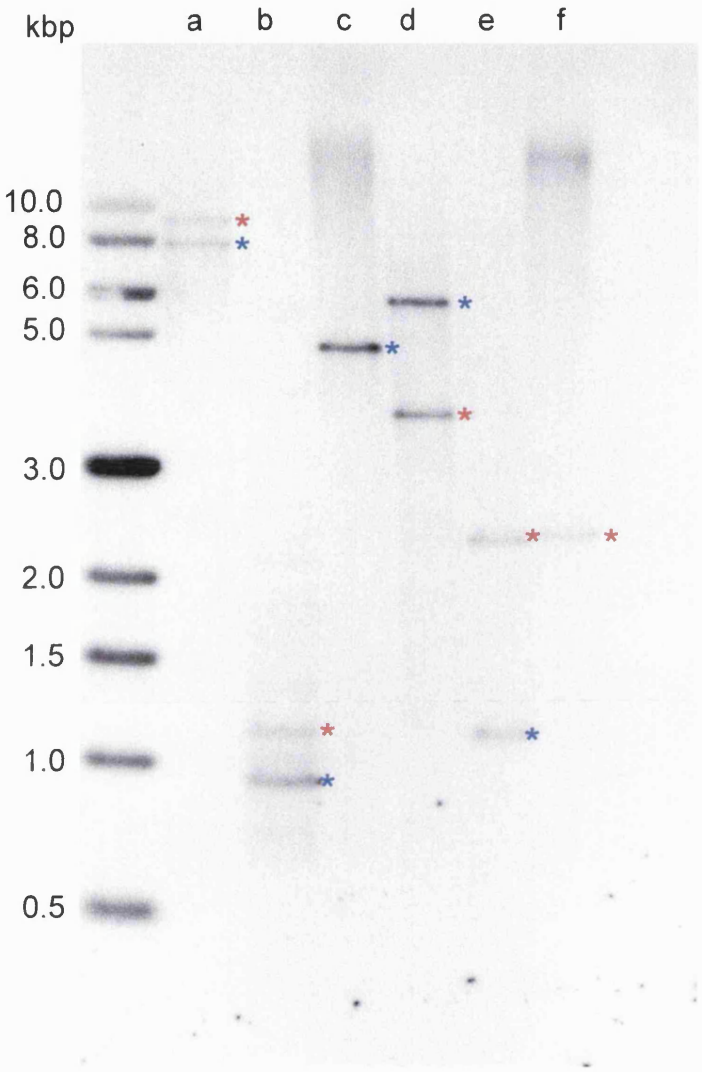
Genomic DNA was prepared from WT *C. reinhardtii*, digested with a) *Sma*I b) *Pvu*II c) *Hind*III d) *Sph*I e) *Pst*I and f) *Hpa*I. The filter was probed with the 1.4kbp *Eco*RI-*Kpn*I fragment from the AV390342 EST clone.

Fig. 4.6 Southern analysis of genomic DNA with the AV390342 EST clone

resulting autoradiogram. The *rpo*D gene is present as a single copy in the nuclear genome of *C. reinhardtii*. Even under low stringency wash conditions a gene family of σ factors cannot be identified. In order to confirm that AV392956 contains the same cDNA as AV390342 plus an additional cDNA the blot was stripped and re-probed with a radiolabelled probe made from the 1.5kbp *Eco*RI-*Xho*I doublet from AV392956. Fig. 4.7 shows the resulting autoradiogram. The bands corresponding to the *rpo*D gene are present again with an additional band in each lane corresponding to the gene of unknown function.

4.2.5 Northern analysis of rpo*D*

In order to investigate the expression patterns of *rpo*D under different growth conditions and to size the transcript a northern blot was carried out. Total RNA was extracted from *C. reinhardtii* WT CC-1021 cells grown in i) TAP + light (mixotrophic growth) ii) TAP – light (heterotrophic growth) and iii) HSM + light (phototrophic growth). The RNA was run on a 1% agarose RNA gel and transferred to nylon by the northern blotting technique. A radiolabelled probe was made from the 1.6kbp *Eco*RI-*Xho*I fragment from the pSKrpoD25′ plasmid, containing the *rpo*D cDNA and used to probe the blot. The resulting autoradiogram is shown in fig. 4.8. A band corresponding to approximately 2.9kb can be seen in the lane containing RNA extracted from cells grown in HSM + light. It is also detectable in the TAP – light lane (albeit at a lower abundance) but is not detectable in the TAP + light RNA. The expression of genes encoding PEP σ factors in both plants and red algae have been shown to be up regulated in response to light (Liu & Troxler, 1996), (Isono *et al.*, 1997). The results of the northern analysis suggest that the *C. reinhardtii* *rpo*D gene is also subject to the same regulation, although this effect may be suppressed by the presence of the acetate carbon source in the medium. Acetate effects have been reported recently by (Heifetz *et al.*, 2000). These findings suggest a role for the *rpo*D product in phototrophic growth, however further work is needed to clarify this.



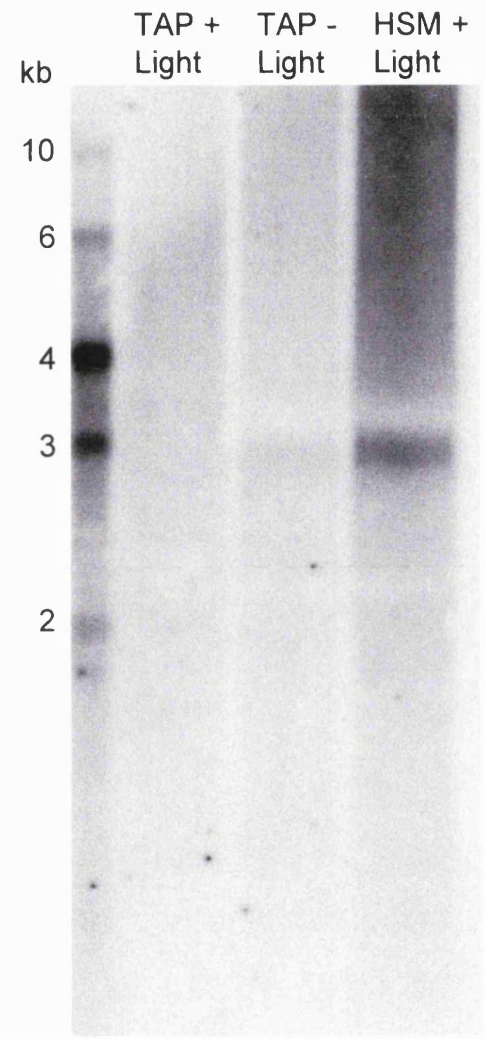
Genomic DNA was prepared from WT *C. reinhardtii*, digested with a) *Sma*I b) *Pvu*II c) *Hind*III d) *Sph*I e) *Pst*I and f) *Hpa*I. The filter was probed with the 1.6kbp *Eco*RI-*Xho*I doublet from the AV392956 EST clone.

Key

* *rpo*D

* Gene of unknown function

Fig. 4.7 Southern analysis of genomic DNA with the AV392956 EST clone



Total RNA was prepared from *C. reinhardtii* WT CC-1021 strain grown in a) TAP + Light, b) TAP - Light c) HSM + Light. 10μg of RNA was run in each lane. The filter was probed with the 1.6kbp *Eco*RI-*Xho*I fragment from the pSKrpoD25' plasmid containing the *rpo*D EST.

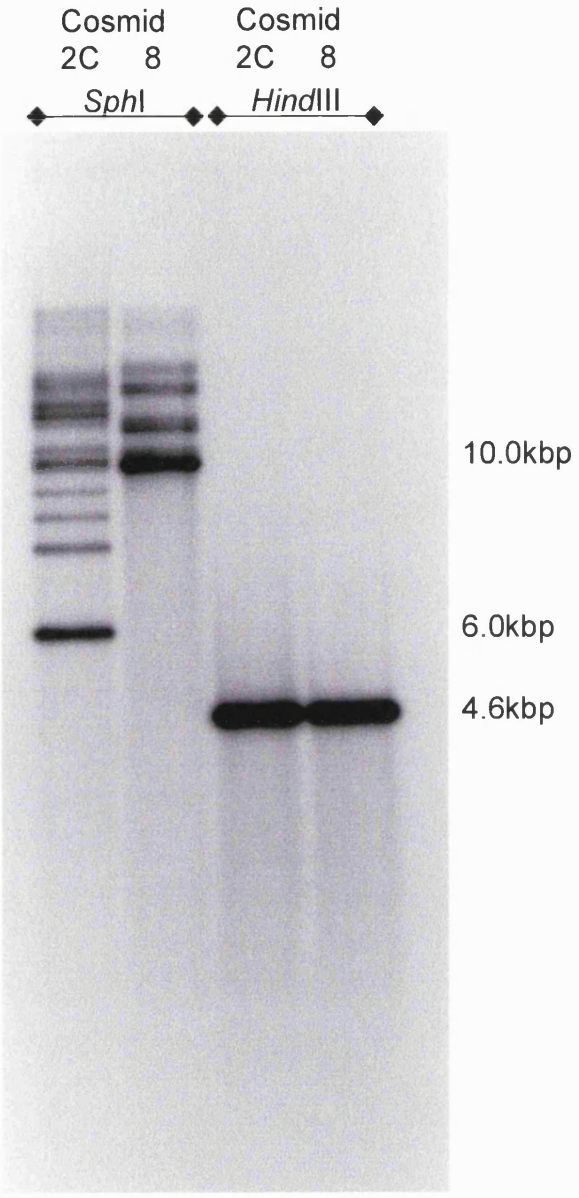
Fig. 4.8 Northern analysis of *rpo*D

4.2.6 Screening a genomic cosmid library

In order to isolate a genomic clone of *rpo*D a *C. reinhardtii* genomic library was screened with a radiolabelled probe made from the 1.4kbp *Eco*RI-*Kpn*I fragment from AV390342. Ten positive clones were obtained in the primary screen and a secondary screen was carried out in order to ensure that the clones were true positives. Two positive clones were identified (cosmids 2C and 8). In order to determine whether the clones represent distinct cosmids from the amplified library the cosmid DNA was extracted from both clones and digested with *Sph*I and *Hind*III. Analysis of the banding pattern of the two cosmids on the agarose gel revealed that the two clones share common restriction fragments, but also have unique fragments (data not shown). This suggests that the genomic clones are distinct, but overlapping clones. The DNA was then transferred to nylon by the Southern blotting technique. The radiolabelled probe made from the 1.4kbp *Eco*RI-*Kpn*I fragment from AV390342 was used to probe the blot. Fig. 4.9 shows the resulting autoradiogram. The *rpo*D gene is located on a 4.6kbp *Hind*III fragment that is common to both cosmids. However, the gene maps to different size *Sph*I fragments (6kbp and 10kbp) in the two clones, confirming that the two clones are overlapping and that a genomic *Sph*I site is beyond the cloned region in one of the cosmids.

4.2.7 Subcloning the rpo*D genomic sequence*

The *rpo*D genomic sequence was subcloned from the cosmid to give a smaller fragment for sequencing. From the genomic Southern blot (fig. 4.6) it appears that the *rpo*D gene lies on a 4.6kbp *Hind*III fragment. The 4.6kbp *Hind*III fragment from cosmid 8 was subcloned into the *Hind*III site of pBluescript SK- to give the pSKrpoD(g) plasmid. This fragment was sequenced as detailed in 4.2.8.



Cosmid DNA was prepared from the *rpo*D cosmids 2C and 8, digested with a) *Sph*I and b) *Hind*III. The filter was probed with the 1.4kbp *Eco*RI-*Kpn*I fragment from the AV390342 EST clone. The faint bands in the *Sph*I lanes are due to incomplete digestion of the DNA.

Fig. 4.9 Southern analysis of the *rpo*D cosmids

4.2.8 Sequencing the rpoD genomic clone

The 4.6kbp *Hind*III fragment of pSKrpoD(g) was sequenced using the T3 and T7 primers, together with internal primers (as detailed in Appendix I). An alignment of the sequence with the known cDNA sequence identified the coding region and two introns. The 5′ end of the *rpo*D gene appears to extend beyond the *Hind*III site, so the complete sequence of the gene was not determined. The 5′ region of the coding sequence and the 5′ UTR is not known. A physical map of the fragment showing the position of the gene on the *Hind*III fragment and the position of the known introns is shown in fig. 4.10.

4.2.9 Analysis of the rpo*D nucleotide sequence*

In order to identify introns in the 5′ region of the gene for which there is no corresponding cDNA sequence the sequence was run through the GeneMark program (http://dixie.biology.gatech.edu/GeneMark/eukhmm.cgi#prot/). This program has been optimised for *C. reinhardtii* nuclear genes using known intron boundary sequences and codon usage data. However, the program failed to identify any introns in this region. The possibility remains, however, that introns are present which have not been identified so a complete cDNA sequence would be necessary to confirm the absence of introns in this region of the gene. This could be achieved by RT-PCR using customised primers to amplify the cDNA sequence from total RNA.

The *Hind*III fragment has a GC content of 63%, which is consistent with other *C. reinhardtii* nuclear genes. The first intron is 167bp and has the consensus sequences of GTGAG at the 5′ splice site and CAG at the 3′ splice site. The second intron is 199bp and has the consensus sequence of GTGCG at the 5′ splice site. The 3′ splice site of the second intron has the TAG sequence, which is only occasionally found in *C. reinhardtii* introns. *C. reinhardtii* introns are an average size of 219bp, but they range from 57bp to 1318bp. In

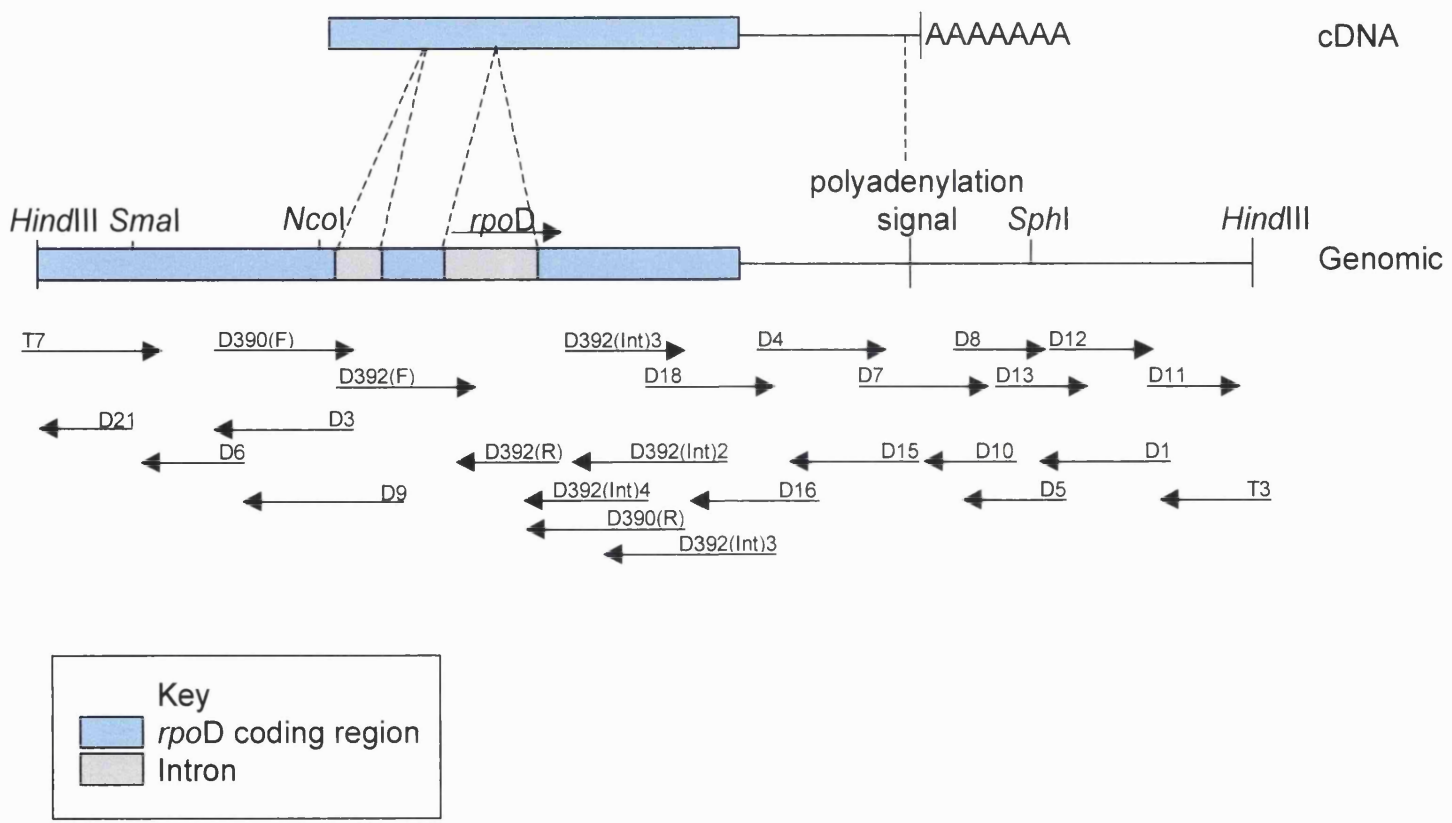


Fig. 4.10 Physical map of the 5kbp region containing the *rpo*D genomic sequence and the sequencing strategy

general the 5′ splice sites conform to the consensus sequence of GTNNG and the 3′ splice sites to the NAG sequence. The stop codon of the *rpo*D gene is TAA, which is the most commonly used stop codon in *C. reinhardtii* nuclear genes. The polyadenylation signal, which conforms to the consensus of TGTAA is 657bp from the stop codon. The polyadenylation signal is 10bp upstream of the start of the polyA tail, which is also a typical characteristic of *C. reinhardtii* nuclear genes (Silflow, 1998).

Table 4.2 A summary of the characteristics of the rpo*D gene*

	*rpo*D	Typical *C. reinhardtii* gene
GC Content	63%	62%
Size of Intron 1	167bp	57-1318bp (average 219bp)
Consensus sequence at the 5′ splice site of intron 1	GTGAG	GTNNG
Consensus sequence at the 3′ splice site of intron 1	CAG	NAG
Size of Intron 2	199bp	57-1318bp (average 219bp)
Consensus sequence at the 5′ splice site of intron 2	GTGCG	GTNNG
Consensus sequence at the 3′ splice site of intron 2	TAG	NAG
Stop codon	TAA	TAA
Polyadenylation signal	TGTAA	TGTAA
Distance of polyadenylation signal from stop codon	657 bases	several hundred bases
Distance of polyA tail from polyadenylation signal	10 bases	10-20 bases

The sequence downstream of the *rpo*D gene was run through the tRNAscan-SE program (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/) to identify any tRNA genes that may be present in this region. No tRNA genes were identified.

From the known *rpo*D coding sequence, deduced from the known genomic sequence the transcript is at least 2691b. The northern blot detected a transcript of ~2.9kb, so this is consistent with the approximate size of the transcript from the sequencing, and does suggest that the 5′ end of the gene extends beyond the cloned region.

4.2.10 Analysis of the RpoD deduced protein sequence

The RpoD protein is at least 678 amino acids with a predicted molecular mass of 72593 Da (73 kDa). The protein is approximately of neutral charge with at least 92 basic and 96 acidic amino acids. The protein has an isoelectric point of 6.351 and a charge of –2.892 at pH 7.0. The codon usage follows that of typical *C. reinhardtii* genes in that most codons use G or C as the nucleotide in the third position. For example arginine is most commonly encoded by the codons CGC or CGG, rather than AGA, AGG, CGA or CGU (fig. 4.11). The amino acid composition of the protein is unremarkable with no notable bias towards any class of amino acid. In terms of secondary structure the protein is predicted to be mainly composed of α helix; there is little β sheet (fig. 4.12).

In order to identify motifs in the RpoD protein, the sequence was run through the PROSITE program (http://expasy.cbr.nrc.ca/cgi-bin/scanprosite/). A σ^{70} motif was identified in the protein sequence (fig. 4.13), thereby classifying this protein as a member of the σ^{70} class of σ factors. If the protein is indeed a component of the chloroplast PEP it should possess a chloroplast targeting sequence at its *N*-terminus (Hiejne *et al.*, 1989). In order to identify a chloroplast transit sequence the sequence was run through the ChloroP program (http://www.cbs.dtu.dk/services/ChloroP/). No chloroplast transit sequence was predicted by this method, but this prediction program is based on chloroplast transit sequences from higher plants and the *C. reinhardtii* sequences differ significantly from these. For example the *C. reinhardtii* chloroplast targeting peptides are more similar to mitochondrial targeting

gca	A	17	aaa	K	2
gcc	A	19	aag	K	36
gcg	A	45	aug	M	16
gcu	A	14			
			uuc	F	8
aga	R	1	uuu	F	4
agg	R	2			
cga	R	2	cca	P	5
cgc	R	28	ccc	P	22
cgg	R	16	ccg	P	13
cgu	R	5	ccu	P	4
aac	N	10	agc	S	22
aau	N	0	agu	S	2
			uca	S	8
gac	D	33	ucc	S	16
gau	D	4	ucg	S	11
			ucu	S	6
ugc	C	3			
ugu	C	0	uaa	*	1
			uag	*	0
caa	Q	1	uga	*	0
cag	Q	17			
			aca	T	7
gaa	E	3	acc	T	8
gag	E	56	acg	T	28
			acu	T	2
gga	G	1			
ggc	G	27	ugg	W	4
ggg	G	9			
ggu	G	5	uac	Y	4
			uau	Y	4
cac	H	6			
cau	H	1	gua	V	1
			guc	V	8
aua	I	0	gug	V	27
auc	I	8	guu	V	4
auu	I	5			
cua	L	2			
cuc	L	10			
cug	L	48			
cuu	L	2			
uua	L	1			
uug	L	5			

Fig. 4.11 Codon usage of the *rpo*D gene. The table shows the number of times a codon appears in the *rpo*D gene

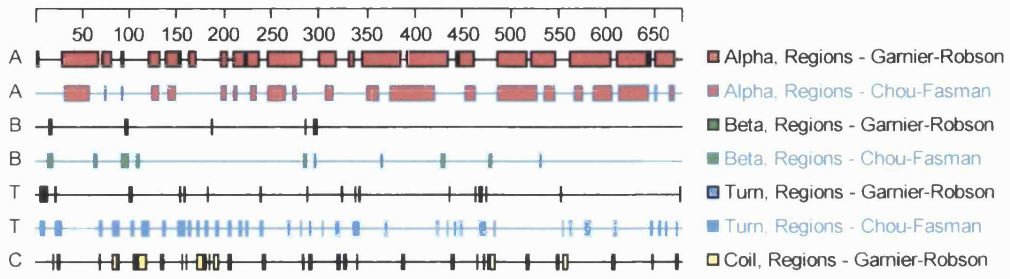


Fig. 4.12 Analysis of the predicted secondary structure of the RpoD deduced protein sequence

Fig. 4.13 Alignment of the amino acid sequences of *C. reinhardtii* RpoD with Sig1 of mustard (*Sinapis alba*) to show the conserved domains. The σ^{70} motif is shown in blue text.

peptides than to higher plant chloroplast targeting peptides in terms of amino acid composition and length. They also have an amphiphilic α helix in common with mitochondrial presequences (Franzén, 1995). Also, since it appears that the RpoD protein sequence is truncated at the *N*-terminus it is probable that part of the targeting sequence is missing. Nonetheless, the sequence was run through the SignalP programe (http://www.cbs.dtu.dk/services/SignalP) in order to identify any possible targeting sequences. Again, this program predicted no targeting sequence.

Alignments of the RpoD protein sequence with the *Synechococcus* SigA and the *Anabaena* SigA proteins show that the RpoD protein has an *N*-terminal extension compared to the cyanobacterial protein (fig. 4.14). This may represent a targeting sequence. A BLAST search (http://www.ncbi.nih.gov:80/BLAST/) using the RpoD protein sequence shows that the *C*-terminal region of the protein shows homology to σ factors from cyanobacteria, bacteria, red algae and higher plants, but the *N*-terminal region does not show significant homology to anything else. Analysis of the proteins from the eukaryotes (red alga, plants and *C. reinhardtii*) shows that they all have *N*-terminal extensions relative to the prokaryotic sequences (fig. 4.14). Again this provides evidence for a chloroplast targeting sequence at the N-terminus of RpoD.

4.2.11 Southern analysis of WT C. reinhardtii *genomic DNA using the pea* rpo*A gene*

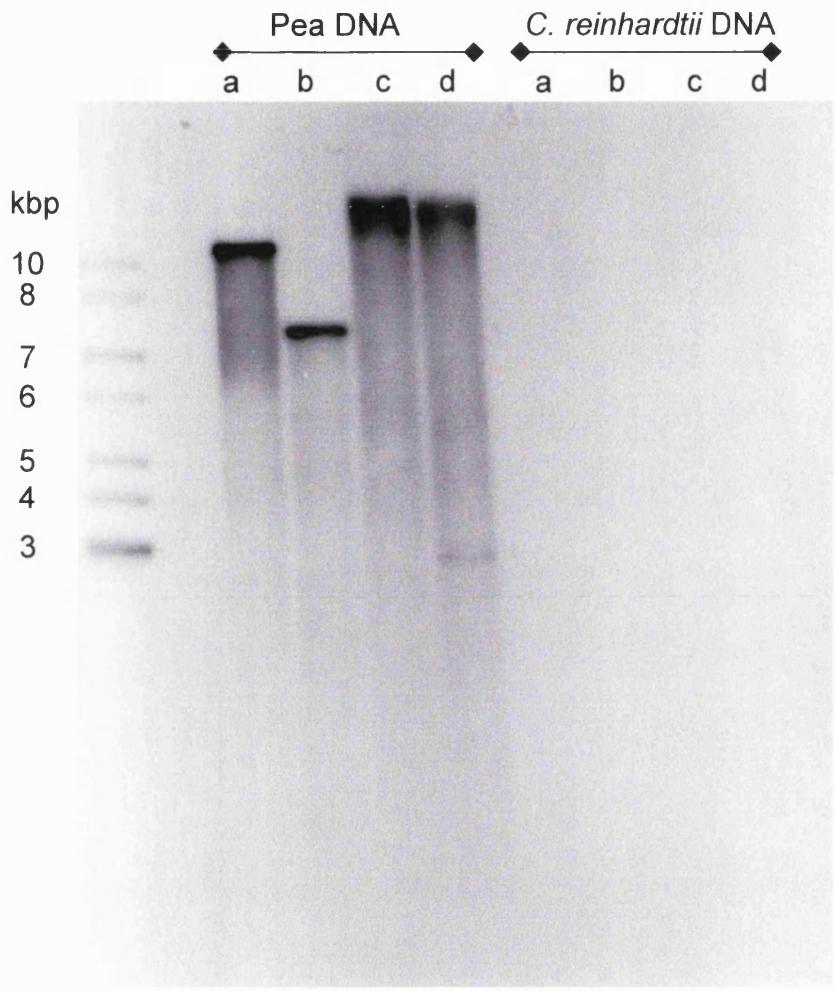
The α subunit of the multi-subunit RNA polymerase of bacteria and plastids is a core component of the enzyme. In most algae and higher plants the α subunit is encoded by the chloroplast gene *rpo*A. Sequencing of the *C. reinhardtii* chloroplast genome has failed to identify an *rpo*A gene. It is therefore likely that this gene has been transferred to the nucleus as discussed in the introduction. In order to identify the *rpo*A gene in total cellular DNA of *C. reinhardtii* a Southern blot was carried out. If the gene

Fig. 4.14 Alignment of the amino acid sequences of *C. reinhardtii* RpoD with σ factors from cyanobacteria, a red alga and higher plants

could be identified using a heterologous probe from pea, this probe could also be used to screen a genomic library. Total DNA was extracted from WT *C. reinhardtii* CC-1021 and pea seedlings and digested with i) *Hind*III ii) *Hpa*I iii) *Pvu*II iv) *Bam*HI. The digested DNA was run on a 1% agarose gel and transferred to nylon by the Southern blotting technique. A radiolabelled probe was made from the 1.0kbp *Hind*III-*Eco*RI fragment from pPscT920, containing the pea *rpo*A gene (Purton & Gray, 1989) and used to probe the blot. Fig. 4.15 shows the resulting autoradiogram. The probe identifies bands of the expected sizes in the lanes containing pea DNA, but no bands can be identified in the lanes containing *C. reinhardtii* DNA, even under low stringency washing conditions. The *rpo*A gene is poorly conserved amongst different species, so if the gene is present in the nuclear genome of *C. reinhardtii* it is probably highly divergent from the chloroplast encoded pea *rpo*A gene. In addition *C. reinhardtii* nuclear genes are GC rich, whereas higher plant chloroplast genes are AT rich, thereby further compromising the heterologous hybridisation.

4.2.12 An attempt to clone rpo*A by RACE*

In order to try to clone *rpo*A from the nuclear genome of *C. reinhardtii* RACE (Rapid Amplification of cDNA Ends) was again used. Degenerate primers were designed to conserved regions of known plant, algal and bacterial *rpo*A genes (fig. 4.16). PolyA mRNA was extracted from WT *C. reinhardtii* CC-1021 cells grown in TAP and converted to cDNA using the oligodT-58 primer and reverse transcriptase. A PCR amplification was carried out using the cDNA template and the RACE1 and rpoA(F) primers. The RACE product was then used as a template in a second amplification using the RACE2 primer, which is nested to the RACE1 primer, and the rpoA(N) primer, which is nested to the rpoA(F) primer. When the final RACE product was run on an agarose gel no band could be detected. Again, this could be due to poor primer design or a lack of *rpo*A expression under the growth conditions used.



Genomic DNA was prepared from pea seedlings and WT *C. reinhardtii* cells, digested with a) *Hind*III b) *Bam*HI c) *Hpa*I and d) *Pvu*II. The filter was probed with the 1.0kbp *Hind*III-*Eco*RI fragment from pPscT920 containing the pea *rpo*A coding sequence

Fig. 4.15 Southern analysis of pea and *C. reinhardtii* DNA to detect the *rpo*A gene

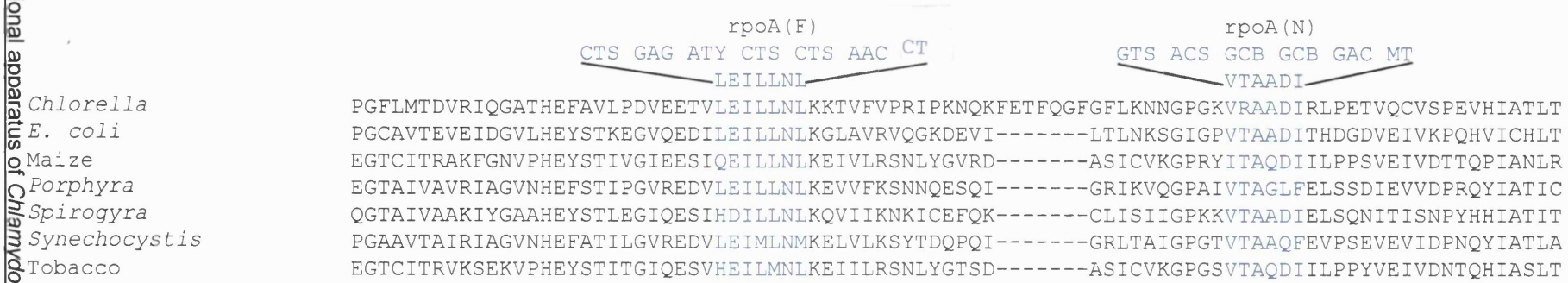


Fig. 4.16 Alignment of portions of the protein sequences of *rpo*A genes from a range of species to show the location of the degenerate primers used in RACE

4.2.13 Searching the EST database for the rpoA gene

In order to identify an EST corresponding to the *rpo*A gene a BLAST search of the EST database (http://www.ncbi.nih.gov:80/BLAST/) was carried out using both the *Synechocystis* and tobacco *rpo*A genes. No *C. reinhardtii* ESTs with significant homology to these genes were identified.

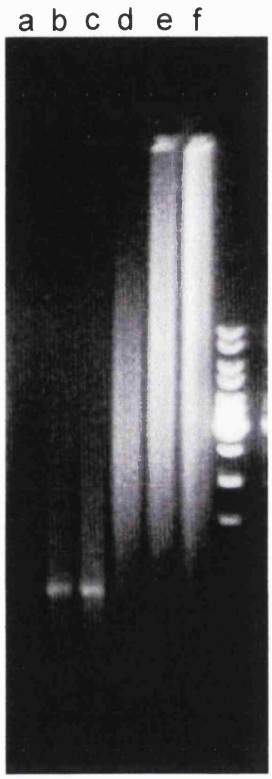
4.2.14 Attempt to clone NEP by PCR

As discussed in the introduction, *C. reinhardtii* is predicted to possess a mitochondrial NEP and the alga may or may not also possess a plastid NEP. In order to attempt to clone a NEP gene from the nuclear genome of *C. reinhardtii* a PCR approach was used. A series of degenerate primers were designed to conserved regions of plant nuclear encoded mitochondrial and chloroplast RNA polymerases, yeast nuclear encoded mitochondrial RNA polymerases, and bacteriophage SP6 and T7 RNA polymerases (fig. 4.17). Combinations of these primers were used in PCR amplifications from WT *C. reinhardtii* total DNA. When the NEPCA(F) and NEPVT(R) primers were used together a band of 1kbp was seen (fig. 4.18). This PCR product was cloned into pBluescript (SK) to give the plasmids pSKNEP2, pSKNEP3, pSKNEP4 and pSKNEP5. pSKNEP2 and 4 contain slightly larger inserts than pSKNEP3 and 5, suggesting that they represent distinct PCR products. The plasmids were sequenced with T3 and T7 primers and the sequences used in a BLAST search (BLASTX) (http://www.ncbi.nih.gov:80/BLAST/). The sequences did not show significant homology to single subunit phage-type polymerases in the database. It is likely that these PCR products resulted from non-specific annealing of the primers.

4.2.15 Searching the EST database for a NEP cDNA

In order to identify an EST corresponding to the NEP a BLAST search of the EST database (http://www.ncbi.nih.gov:80/BLAST/) was again carried out

Fig. 4.17 Alignment of portions of the protein sequences of phage-type RNA polymerases from a range of species to show the location of the degenerate primers used in PCR



PCR products from genomic DNA extracted from WT *C. reinhardtii*, amplified with NEPCA(F) and NEPVT(R) primers at different Mg^{2+} concentrations. a) No Vent, b) 2mM $MgSO_4$, c) 3mM $MgSO_4$, d) 4mM $MgSO_4$, e) 5mM $MgSO_4$, f) 6mM $MgSO_4$

Fig. 4.18 Degenerate PCR to clone a NEP

using the *Arabidopsis* gene. No *C. reinhardtii* ESTs with significant homology to this gene were identified.

4.3 Discussion

A gene encoding a putative PEP σ factor has been cloned from the *C. reinhardtii* nuclear genome and partially sequenced. This gene has been named *rpo*D. A partial cDNA clone was obtained via the Japanese EST collection, which showed significant similarity to an *Arabidopsis* σ factor gene and this cDNA was used to screen a genomic library. Two distinct cosmids were identified which contained the corresponding genomic sequence and a 4.6kbp *Hind*III fragment, known to contain the *rpo*D genomic sequence was subcloned and sequenced. The 5′ end of the *rpo*D genomic sequence extends beyond the *Hind*III fragment, so in order to obtain a complete genomic sequence a larger fragment needs to be subcloned from the cosmid and the 5′ end of the gene, including the 5′ UTR, sequenced. Also the cDNA is truncated, and corresponds to approximately half of the complete cDNA. In order to verify the coding sequence and the presence of introns in the 5′ region of the gene RT-PCR must be carried out in order to obtain a complete cDNA sequence.

The *rpo*D gene is present as a single copy in the nuclear genome of *C. reinhardtii*, and a gene family of σ factors could not be identified using the EST clones to probe a Southern blot. Furthermore, only one *rpo*D gene was identified by the EST analysis of more than 11,500 EST clones. A gene family of σ factors were identified in the nuclear genome of *C. caldarium* by low stringency Southern analysis using a cDNA probe (Liu & Troxler, 1996), but this appears not to be the case in *C. reinhardtii*. This could be because *C. reinhardtii* has fewer σ factor genes than the red alga, but is more likely to be due to divergence between the σ factor genes in this species. Previous

biochemical isolation of σ factor proteins from *C. reinhardtii* indicated that at least two of these proteins are present (Surzycki & Shellenbarger, 1976).

A transcript of ~2.9kb was detected by northern analysis in RNA extracted from *C. reinhardtii* WT CC-1021 cells grown in heterotrophic and phototrophic conditions, but not in mixotrophic conditions. The size of 2.9kb is consistent with a deduced transcript of at least 2691b from the genomic sequence. The presence of the transcript in cells grown in phototrophic conditions is consistent with previous findings that σ factor genes are transcribed in response to illumination (Liu & Troxler, 1996), (Isono *et al.*, 1997) and ties in with a predicted role in phototrophic growth. The lack of a transcript in cells grown mixotrophically is difficult to explain, as σ factor gene transcripts are normally present in illuminated cells. Acetate present in the growth medium has been shown to have an effect on CO_2 fixation and O_2 evolution, so it is conceivable that it may suppress transcription of σ factor genes that are involved in phototrophic growth. This is not consistent with the presence of an *rpo*D transcript in cells grown heterotrophically, however.

The *rpo*D gene has all the typical characteristics of a *C. reinhardtii* nuclear gene in terms of GC content (63%), the presence of two introns, the 5′ and 3′ splice sites of which conform to the *C. reinhardtii* consensi of GTNNG at the 5′ splice sites and NAG at the 3′ splice sites. The introns are 167bp and 199bp respectively, close to the average of 219bp for nuclear genes, and well within the range from 57bp to 1318bp. The *rpo*D gene has TAA as a stop codon, which is the most commonly used stop codon in *C. reinhardtii* nuclear genes and the polyadenylation signal also conforms to the consensus of TGTAA. The polyA tail begins 10 bases downstream of this signal, as is typical (Silflow, 1998). The *rpo*D gene has a 3′ UTR of 657bp.

It is difficult to draw firm conclusions regarding the RpoD predicted protein as the complete coding sequence has not been determined, and the protein is

truncated at the *N*-terminus. The protein is, however, at least 72593 Da (73 kDa), and consists of at least 678 amino acids. The protein has at least 92 basic and 96 acidic amino acids and as such is neutral in charge. The protein is unremarkable in terms of amino acid composition and in terms of secondary structure is predicted to be mainly composed of α helix.

No putative chloroplast transit sequence could be identified by either the ChloroP programme or the SignalP program, although amino acid alignments with cyanobacterial σ factors show that the *C. reinhardtii* protein has an N-terminal extension. This may represent a chloroplast transit sequence, which is truncated as a result of the lack of sequence data from the 5′ end of the gene. The ChloroP and SignalP programs are based on transit peptides from non algal species and the *C. reinhardtii* sequences differ significantly from these. For example they are often more similar to mitochondrial targeting peptides than to higher plant chloroplast targeting peptides (Franzén, 1995).

The PROSITE program identified a σ^{70} motif in the RpoD protein sequence. In addition there is strong homology between the *C. reinhardtii* RpoD and σ factors from *Anabaena*, *Synechococcus*, *Arabidopsis*, *C. caldarium* and *S. alba* within the conserved domains identified in prokaryotic σ factors. It is reasonable to assume a role for RpoD as a PEP σ factor, but more work needs to be carried out to prove that the protein is chloroplast located. It would be useful to raise antibodies to the protein sequence in order to carry out immunolocalisation experiments and visualise the protein in the chloroplast. These antibodies could also be used in expression studies to confirm or refute the findings of the northern analysis.

The deduced amino acid sequence of *rpo*D allows an analysis of why the RACE failed to clone this gene. The sigmaF primer was raised to an amino acid sequence of YWWIRQG, the nucleotide sequence being TACTGGTGGATYCGBCAGG. From the deduced amino acid sequence of

*rpo*D it can be seen that the amino acid sequence is HWWIRQP, the nucleotide sequence being CACTGGTGGATCCGCCAGC. The presence of a codon for tyrosine (TAC) rather than histidine (CAC) at the 5′ end of the primer was probably not the cause of the failure as the primers can "hang off" the template at the 5′ end and extension can occur. The presence of the first nucleotide of the codon for a glutamine residue (G) at the 3′ end of the primer was probably the cause of the failure as the *rpo*D sequence has a C at this point. A mis-prime at the 3′ end will prevent extension. In addition the sigmaN primer was raised to an amino acid sequence of RERVRQ (CGBGAGCGBGTSCGBCAG), rather than RERIRQ (CGCGAGCGCATCCG TCAG). A difference such as this in the middle of the primer would be sufficient to cause mis-priming. Also RACE was carried out using mRNA extracted from *C. reinhardtii* cells grown in TAP + light, and from the northern blot it appears that the *rpo*D transcript is of low abundance in cells grown in these conditions.

Differences such as these between the RACE primers used in the reaction to clone a nuclear *rpo*A gene and the actual *rpo*A sequence could explain its failure. The *rpo*A gene is highly divergent in amino acid sequence amongst chloroplast encoded genes, so a nuclear encoded gene would be expected to be very different from these. This also explains why the chloroplast encoded pea *rpo*A gene failed to detect an *rpo*A gene in the *C. reinhardtii* genome by low stringency Southern analysis. The lack of an *rpo*A gene, which encodes an essential PEP subunit, in the chloroplast genome of *C. reinhardtii* along with evidence for the transfer of the *rpo*A gene from the plastid genome of Plasmodium to the nucleus (Sato *et al.*, 2000) suggests that this has also occurred in *C. reinhardtii*. The complete sequencing of both the chloroplast and nuclear genomes of *C. reinhardtii* will eventually answer this question.

Again, the failure of the PCR to clone a NEP gene could be due to these sorts of differences between the primer sequences and the actual sequence

of a *C. reinhardtii* NEP gene. It is complicated in this case as a NEP may not exist in *C. reinhardtii*, so the failure is not surprising. Again, a completely sequenced nuclear genome will allow this question to be answered conclusively.

Chapter 5 – Analysis of the RpoC2 Protein by Epitope Tagging

5.1 Introduction

Epitope tagging of gene products *in vivo* has become a powerful tool in molecular research. A gene product can be made immunoreactive to commercially available monoclonal antibodies by the addition of 6-30 amino acids. The resulting protein can then be used in a large number of immunochemical and immunocytochemical techniques. The traditional method of raising antibodies to the protein of interest can be slow, expensive and unpredictable in terms of immunoreactivity. For example, the protein of interest can be poorly immunogenic or the antibodies may cross-react with related proteins. In addition, epitope tagging has other advantages over the use of traditional antibodies. For example, the commercially available antibodies have been characterised in terms of class, titre, affinity, avidity, background, response to temperature, pH, buffer conditions, etc. Epitope tagging of proteins also allows experiments that would not be possible using traditional methods. For example, the product of a transgene can be distinguished from the endogenous gene (Jarvik & Telmer, 1998).

A number of epitope-antibody combinations are available for epitope tagging (Table 5.1).

The 6x-histidine epitope has an additional property that is useful in protein analysis. Histidine is a strong metal-coordinating residue, and when sufficiently large numbers of these residues are accessible the protein can be purified from crude mixtures by immobilized metal-affinity chromatography (IMAC) (Sulkowski, 1985), (Arnold & Haymore, 1991) (fig. 5.1). Engineering of 6x-histidine residues into an external, non-functional portion of a given protein (the *N*- or *C*-termini are normally used as they are often on the outside of a folded polypeptide, and are rarely included in the active site (Jarvik & Telmer, 1998)) provides a useful method for purifying the protein. A large number of 6x-histidine tagged proteins have been successfully isolated from a range of species, but in particular 6x-histidine tagged PSII has been

Table 5.1 Commercially available epitope / antibody combinations (Jarvik & Telmer, 1998).

Epitope	Size	Sequence	Antibody	Immunogen
AU1	6 aa	DTYRYI	AU1	BPV-1
AU5	6 aa	TDFYLK	AU5	BPV-1
B Tag	6 aa	QYPALT	D11, F10	VP7 of bluetongue viruses
c-myc	10 aa	EQKLISEEDL	9E10	Synthetic peptide: residues 408-439 of human c-myc gene product
FLAG	8 aa	DYKDDDDK	M1, M2, M5	Synthetic peptide containing enterokinase cleavage site
Glu-Glu	6 aa	EYMPME or EFMPME	GLU-GLU	Synthetic peptide EEEEYMPME of polyoma virus medium T antigen
HA	9 aa	YPYDVPDYA	12CA5, 3F10, HA.1	Human influenza virus hemagglutinin, HA1
His6	6 aa	HHHHHH	6-His, 6xHis, HIS-11	Recombinant His-tagged fusion protein
HSV	11 aa	QPELAPEDPED	HSV•TAG®	Synthetic peptide from HSV glycoprotein D
HTTPHH	6 aa	HTTPHH	PHHTT	Synthetic peptide PHHTTPHHTTPHHTT
IRS	5 aa	RYIRS	IRS1	Synthetic peptide
KT3	6 aa	PPEPET	KT3	SV40 large T antigen
Protein C	12 aa	EDQVDPRLIDGK	HPC4	Human Protein C
T7	11 aa	MASMTGGQQMG	T7•Tag®	Leader peptide of phage T7 major capsid protein
V5	14 aa	GKPIPNPLLGLDST	V5	P/V of paramoxyvirus SV5
VSV-G	6 aa	MNRLGK	P5D4	Synthetic peptide: C-terminal 15 residues of vesicular stomatitis virus glycoprotein

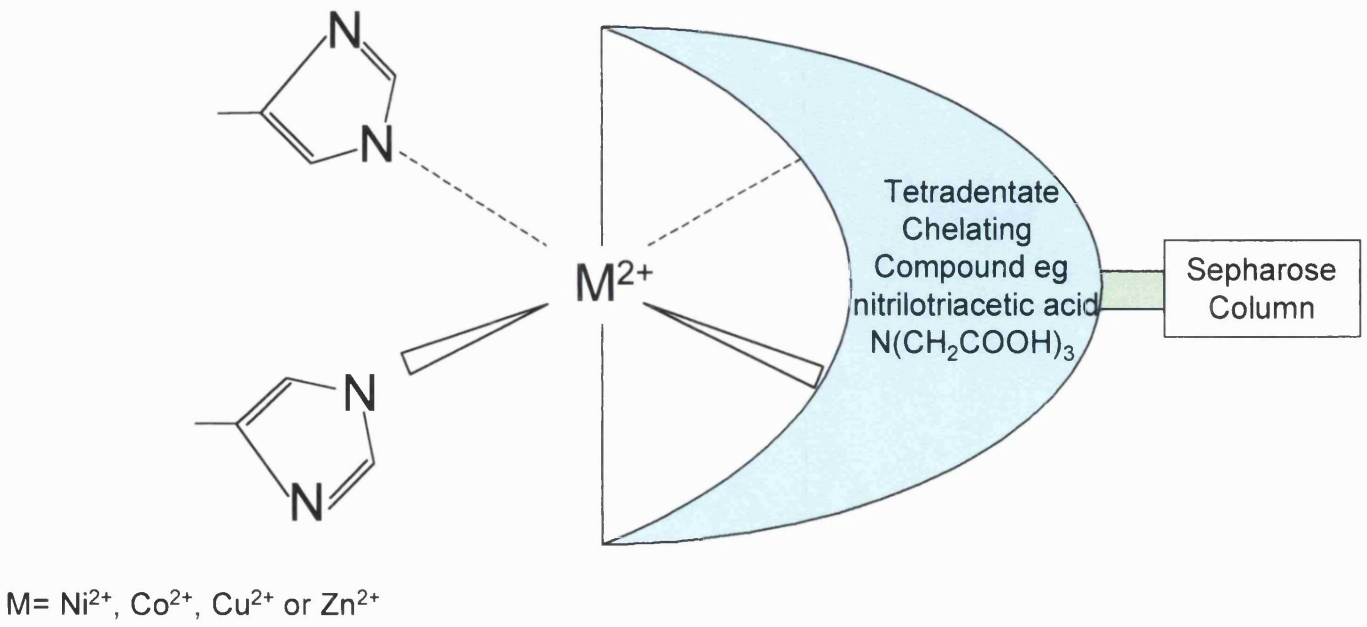


Fig. 5.1 Binding of a 6x-histidine tag to a metal column

isolated from *C. reinhardtii* chloroplasts (Sugiura *et al.*, 1998) and tagging of the β′ subunit of the *E. coli* RNA polymerase allowed the isolation of a transcriptionally active complex (Kashlev *et al.*, 1993). These experiments provide precedents for the 6x-histidine tagging of a subunit of PEP and subsequent isolation from *C. reinhardtii* chloroplasts.

The *rpo*C2 gene of *C. reinhardtii* encoding for the PEP β″ subunit is unusually large and the deduced amino-acid sequence shows blocks of homology to tobacco *rpo*C2. It is possible that this protein is being processed at the post-translational level (protein splicing) as putative inteins have been identified in other *C. reinhardtii* chloroplast genes (Wang & Liu, 1997) as discussed in 1.3.3. The RpoC2 protein of *C. reinhardtii* is an ideal candidate for 6x-histidine tagging as it would allow the protein to be sized, and the question of protein splicing to be addressed. Commercially available anti-His monoclonal antibodies could be used in a western blot to detect the protein, and allow it to be sized or the 6x-histidine tagged protein could be purified by IMAC. An advantage of this approach to purification is that a direct link can be made between the protein that is purified and the gene(s) encoding that protein. This had previously been a stumbling block in chloroplast RNA polymerase purifications as discussed in 1.4.2.

The haemagglutinin (HA) epitope has been successfully used in a range of immunochemical and immunocytochemical techniques. For example in the budding yeast (*Saccharomyces cerevisiae*) an HA tagged protein (Rna15) was efficiently immunoprecipitated (Cullin & Minvielle-Sebastia, 1994). The HA epitope has been used to tag nuclear genes of *C. reinhardtii*. Kozminski and co-workers tagged an α-tubulin gene and used anti-HA antibodies in immunofluorescence and western blot analyses (Kozminski *et al.*, 1993). A *C. reinhardtii* nuclear gene which is chloroplast targeted (*Nac2*) was also 3x-HA tagged and anti-HA antibodies were used in western analysis (Boudreau *et al.*, 2000). No non-specific binding of the anti-HA antibodies to *C. reinhardtii*

proteins was observed. The chloroplast genome of *C. reinhardtii* has a specific codon bias. The existing forms of the HA epitope do not match this and may not be efficiently expressed. In order to use the HA epitope in *C. reinhardtii* chloroplasts an epitope which is codon optimised for chloroplast expression must be synthesised. A cassette with three tandem repeats of the epitope sequence has been reported to increase the avidity of the antibody for the epitope (Tyers et al., 1992). It may be preferable to synthesise a codon optimised triple HA epitope cassette.

A double epitope cassette, consisting of a 3x-HA epitope with a 6x-histidine tag would be another useful tool allowing isolation of chloroplast proteins by IMAC and immunodetection using anti-HA antibodies.

The aims of this chapter are;

- To tag the RpoC2 protein with 6x-histidine residues and use western analysis with anti-his antibodies or IMAC followed by SDS-PAGE to size the protein.

- To synthesise a 3x-HA epitope cassette that is codon optimised for chloroplast expression and test its immunoreactivity to anti-HA antibodies.

- To synthesise a 6x-histidine tagged 3x-HA double epitope cassette to allow detection and isolation of chloroplast proteins.

5.2 Results

5.2.1 Construction of the 6x-histidine tagged rpo*C2 gene*

In order to tag the RpoC2 protein with 6x-histidine residues to allow subsequent western analysis or IMAC a PCR based strategy was designed to

add six codons coding for 6x-histidine residues onto the 3′ end of the *rpo*C2 gene (the *C*-terminus of the RpoC2 protein). The RpoC2 protein is poorly conserved at the *C*-terminus and is the most suitable site for the 6x-histidine tag. The CAC codon coding for a histidine residue was chosen as this is the codon preferentially used in *C. reinhardtii* chloroplasts. In addition a *Stu*I site was added to the end of the 6x-histidine tag as a molecular marker to follow the tag. The cloning strategy is summarised in fig. 5.2. A 220bp region at the 3′ end of *rpo*C2 was amplified from the pSKEco10 plasmid using the rpoC-PmI and rpoC-his primers. The rpoC-PmI primer and the rpoC-his primer are homologous to the *rpo*C2 sequence, but the rpoC-his primer has a non-homologous tail consisting of six CAC codons, coding for six histidine residues, which is in-frame with the 3′ end of *rpo*C2. This 220bp PCR product was blunt cloned into the *Hinc*II site of the cloning vector pUC9, to give the plasmid pUC9-*rpo*C2-his. In order to select for transformants when the construct is used to transform *C. reinhardtii*, the *aad*A cassette conferring resistance to spectinomycin, was cloned into the *Bam*HI site of pUC9 downstream of, and in the same orientation to, the *rpo*C2 gene and the 6x-histidine tag. This created the pUC9-*rpo*C2-his-*aad*A plasmid.

To create the pSKEco10-his-*aad*A plasmid a 170bp *PmlI*-*Pme*I fragment was removed from the pSKEco10 plasmid. The 2.2kbp *PmlI*-*Sma*I fragment from the pUC9-*rpo*C2-his-*aad*A plasmid containing the 3′ end of *rpo*C2, the 6x-histidine tag and the *aad*A cassette was sub cloned into the *PmlI*-*Pme*I sites of pSKEco10 to be in-frame with the *rpo*C2 gene.

Sequencing of the pSKEco10-his-*aad*A plasmid confirmed that the construct was correct, that no errors had been introduced in PCR and that the 6x-histidine tag was in-frame with the *rpo*C2 gene (fig 5.3).

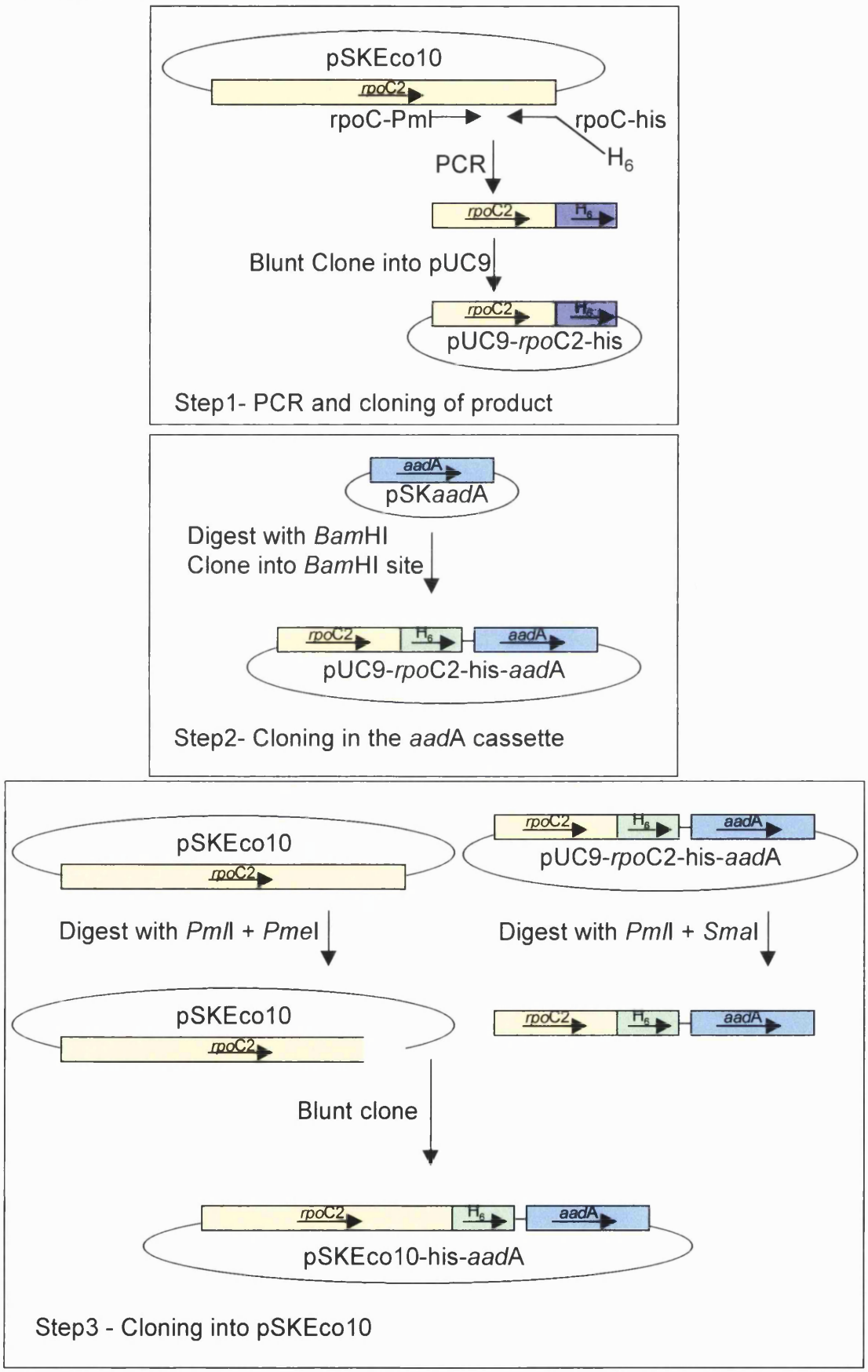


Fig 5.2 Cloning strategy to tag *rpo*C2 with 6x-histidine residues

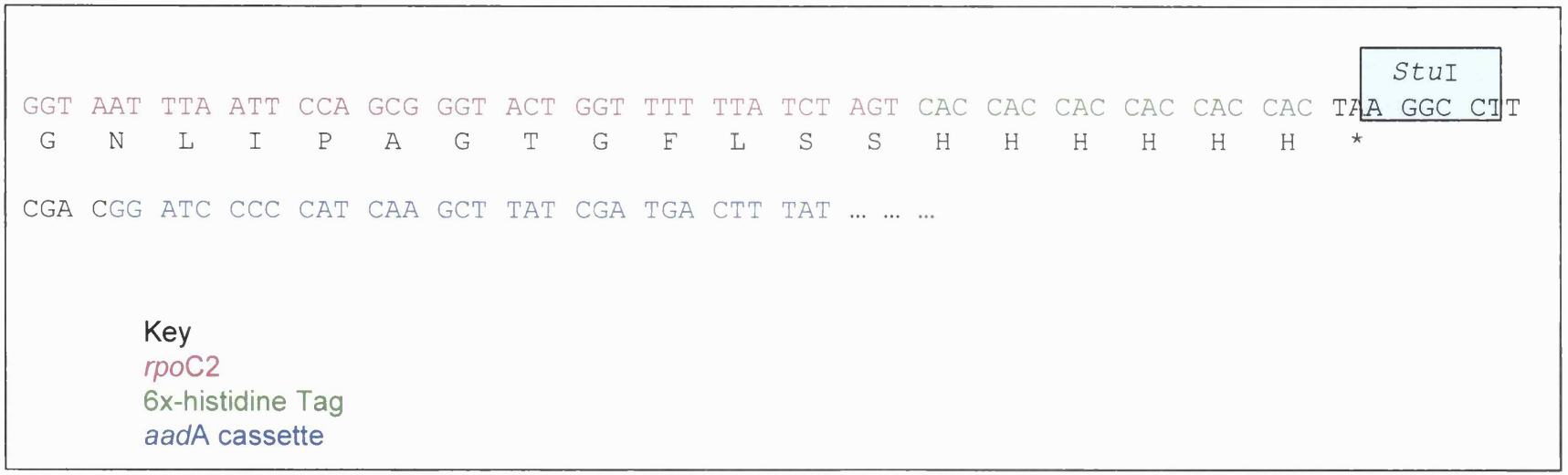


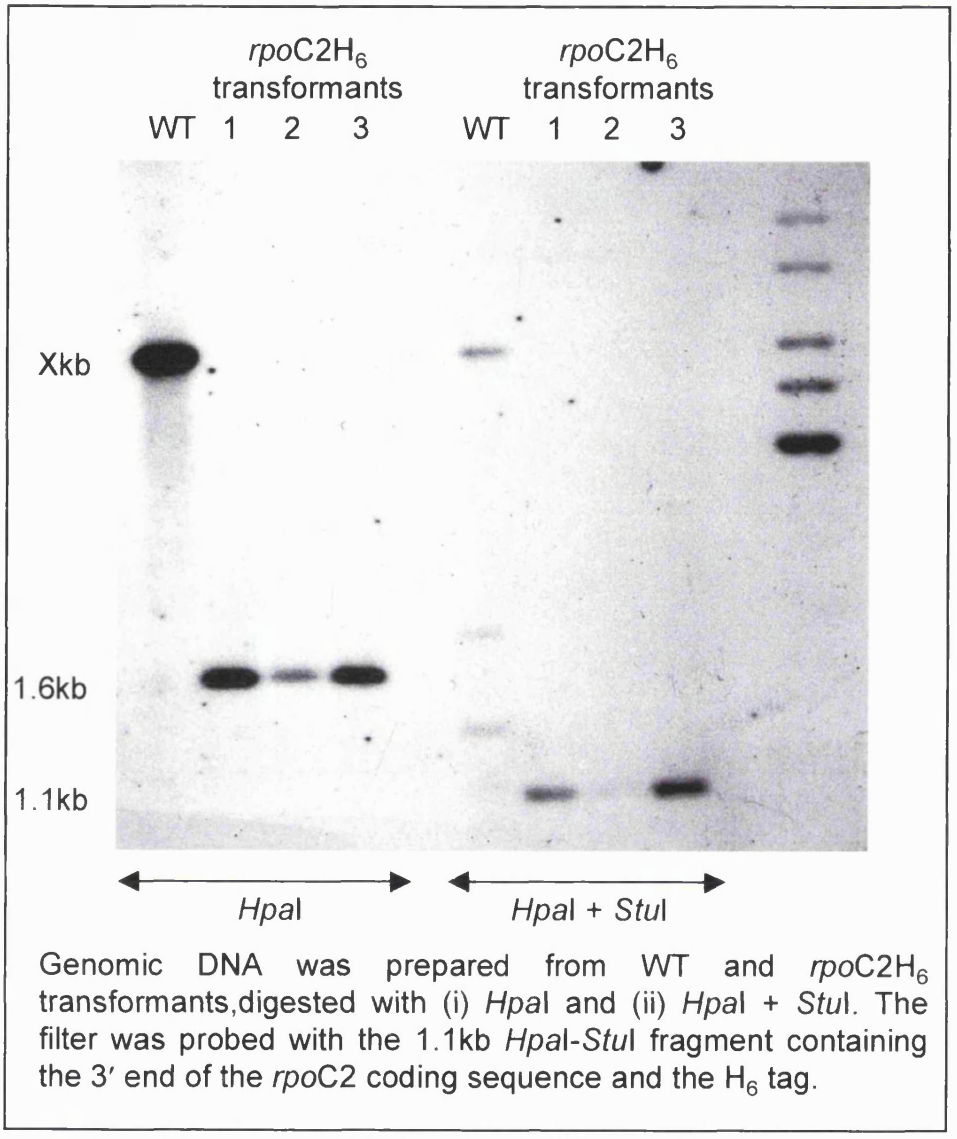
Fig. 5.3 Sequence of the 6x-histidine tagged *rpo*C2 construct

5.2.2 Transformation of C. reinhardtii with the pSKEco10-his-aadA plasmid

The pSKEco10-his-*aad*A plasmid was used to transform WT *C. reinhardtii* CC-1021 cells plated onto TAP plates supplemented with spectinomycin at 100μg/ml (spc^{100}) using the biolistic method. The resulting transformants were taken through three rounds of single colony isolation on TAP + spc^{100} plates to obtain homoplastic transformants. The homoplastic state of the transformants was confirmed by Southern analysis as discussed in 5.2.3.

5.2.3 Southern analysis of the rpo*C2H*$_6$ *transformants*

In order to confirm the presence of the 6x-histidine tag and the *aad*A cassette in the transformants Southern analysis was carried out. The chloroplast contains about 80 copies of its genome per organelle in *C. reinhardtii*, so Southern analysis also determined whether the change had been incorporated into every copy of the genome (homoplastic). Total DNA was extracted from three transformants and WT *C. reinhardtii* and digested with i) *Hpa*I and ii) *Hpa*I + *Stu*I. The digested DNA was run on a 1% agarose gel and transferred to nylon by the Southern blotting technique. A radiolabelled probe was made from the 1.1kbp *Hpa*I-*Stu*I fragment from pSKEco10-his-*aad*A and used to probe the blot. Fig. 5.4 shows the resulting autoradiogram. The transformants show a band of 1.6kbp when digested with *Hpa*I due to the presence of an *Hpa*I site in the *aad*A cassette. This confirms the presence of the *aad*A cassette. The *Hpa*I site in the WT is further downstream, so the band is much larger. A *Stu*I site is present in the 6x-histidine tag so digesting the transformant DNA with *Hpa*I and *Stu*I gives a smaller band of 1.1kbp. This confirms the presence of the 6x-histidine tag. Again, there is no *Stu*I site within this region in the WT, so the band is much larger. The transformants are homoplastic for the change as no WT bands can be detected in the transformant lanes.



Genomic DNA was prepared from WT and *rpo*C2H_6 transformants,digested with (i) *Hpa*I and (ii) *Hpa*I + *Stu*I. The filter was probed with the 1.1kb *Hpa*I-*Stu*I fragment containing the 3′ end of the *rpo*C2 coding sequence and the H_6 tag.

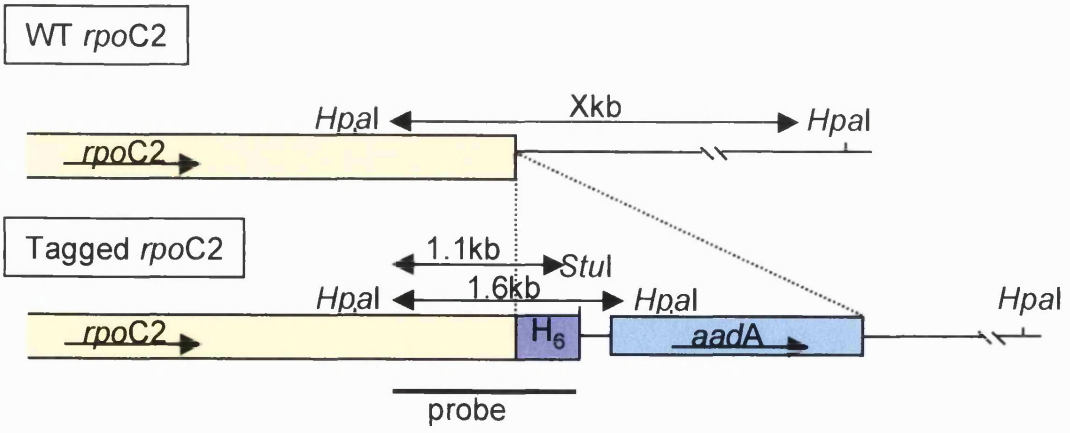


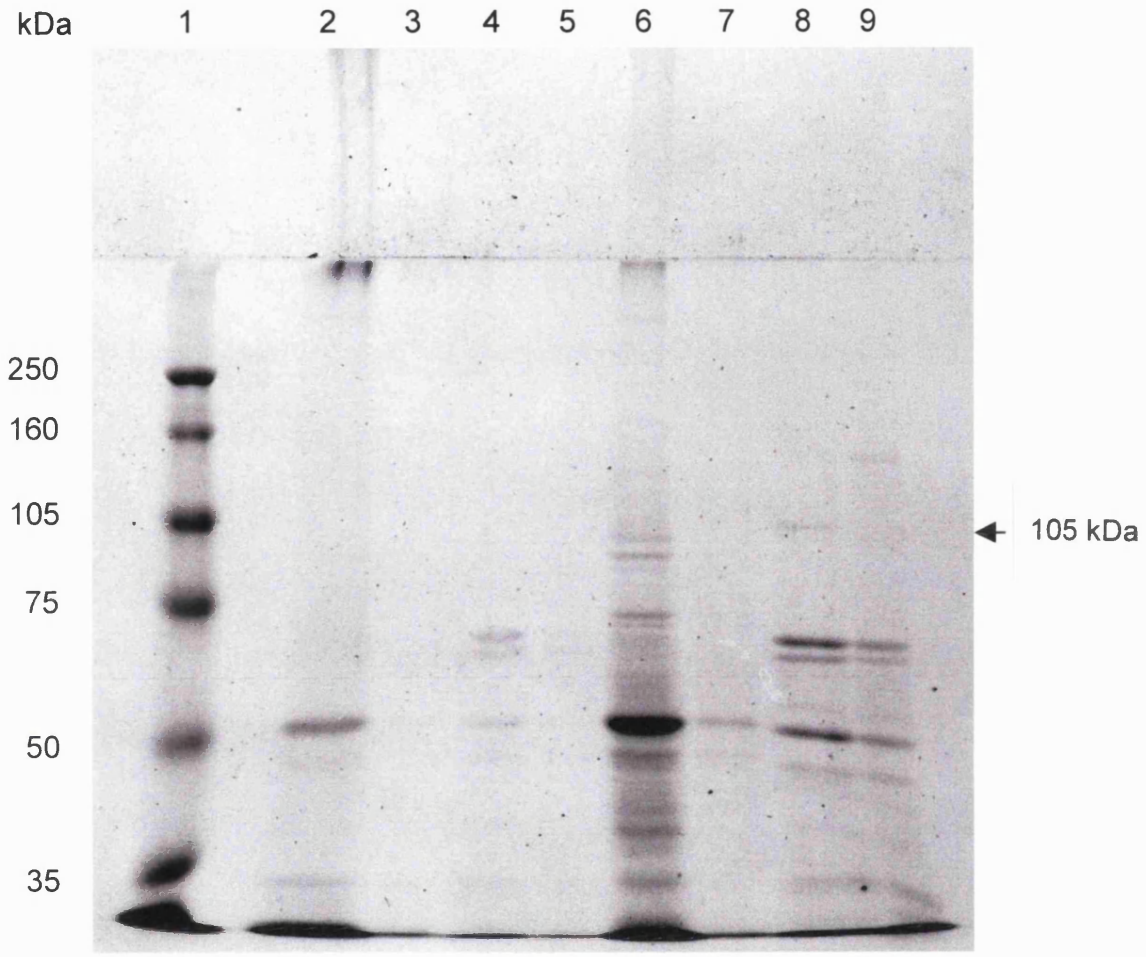
Fig 5.4 Southern analysis of *rpo*C2H_6 transformants

5.2.4 *Western analysis of rpoC2 6x-histidine tagged transformants*

In order to size the RpoC2H$_6$ protein, western analysis using anti-His (C-term) HRP-linked antibodies was carried out on *rpo*C2 6x-histidine tagged transformants. Crude cell lysate, thylakoid membranes, NaCl wash of thylakoid membranes and the soluble fraction were extracted from the *rpo*C2 6x-histidine tagged transformants and used in western analysis. Dot blots using a number of purified 6x-histidine tagged proteins and tetra-his antibodies were also carried out. Despite numerous repetitions of these experiments no cross reactivity of either of the antibodies was seen against any of the samples used. Poor affinity is a known problem of the commercial anti-His antibodies and this could account for the failure of the western analysis.

5.2.5 *Isolation of RpoC2H$_6$ using IMAC*

The affinity of the 6x-histidine tag for Ni^{2+} was exploited in an attempt to purify the RpoC2H$_6$ protein as a single polypeptide, or as part of the PEP enzyme complex. It was hoped that this would allow either identification and sizing of a band corresponding to RpoC2H$_6$, or would produce a sample enriched in PEP / RpoC2H$_6$ which could be used in subsequent western analysis. The purification experiment was carried out as detailed in 2.11.4 from both membrane and soluble cell fractions. The samples were run on a 7.5% SDS-PAGE gel and stained with Coomassie Blue (2.11.1). The results are shown in fig. 5.5. The predicted size of RpoC2, based on its deduced amino acid sequence is around 357.8kDa. No band of this size could be detected. This could be because the protein does not stain with Coomassie Blue, or it could be that the size of RpoC2H$_6$ *in vivo* is smaller than 357.8kDa due to post-translational processing. Although the samples purified from WT and the *rpo*C2H$_6$ mutant share a number of bands, indicating that there is non-specific binding of *C. reinhardtii* proteins to the Ni^{2+} column, a band corresponding to approximately 105kDa is present in the membrane fraction



Lane	Sample
1	Markers
2	rpoC2H_6 Soluble Fraction - Flow through
3	rpoC2H_6 Soluble Fraction - Wash
4	rpoC2H_6 Soluble Fraction - Elution
5	WT Soluble Fraction - Elution
6	rpoC2H_6 Membrane Fraction - Flow through
7	rpoC2H_6 Membrane Fraction - Wash
8	rpoC2H_6 Membrane Fraction - Elution
9	WT Membrane Fraction - Elution

5.5 Coomassie stained SDS-PAGE of protein isolated by Ni^{2+} affinity chromatography

from the *rpo*C2H$_6$ mutant which is absent from the WT. This may represent a post-translationally spliced variant of the RpoC2H$_6$ protein.

Western analysis was not carried out on these samples in the light of the lack of cross reactivity of the anti-His antibodies to other purified 6x-histidine tagged proteins.

5.2.6 Construction of the codon optimised 3x-haemagglutinin (HA) cassette

As a second strategy to detect RpoC2 by western analysis, a 3x-HA epitope cassette that is codon optimised for chloroplast expression was synthesised. The 3x-HA epitope was synthesised by a PCR method that effectively allows a DNA molecule to be synthesised *in vitro* without the need for a template. A summary of this strategy is shown in fig. 5.6. Two large overlapping primers (HA-5′ and HA-3′) were designed which together consist of the nucleotide sequence which encodes for three copies of the HA epitope. For each amino acid in the epitope the codon preferentially used in *C. reinhardtii* chloroplasts was used (taken from Harris, 1989), although some variation was used in the nucleotide sequence to prevent slippage of the primers during annealing.

The HA-5′ and HA-3′ primers were allowed to anneal in a PCR cycle, then Vent DNA polymerase was used to prime from these primers and extend to create a double stranded DNA molecule of 97bp. The double stranded nature of the molecule was confirmed by digestion with *Nco*I as restriction enzymes will only digest double stranded DNA. There is an *Nco*I site downstream of the overlap, so only successfully extended products will cut to give two fragments of 36bp and 61bp.

The PCR product was then used as a template in a conventional PCR using the primers HA(2.1) and HA(2.2) to increase the amount of product and also incorporate *Eco*RI sites to facilitate cloning. This PCR product was then cut

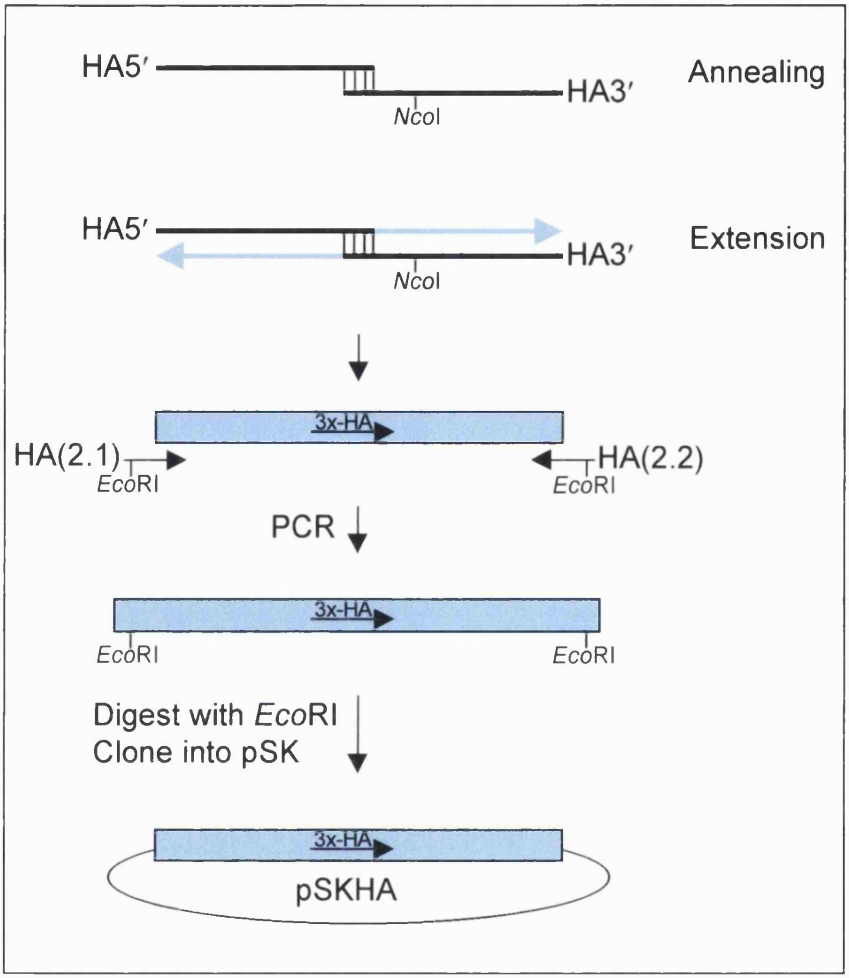


Fig. 5.6 *de novo* synthesis of a codon optimised 3x-HA epitope

with *Eco*RI and cloned into the *Eco*RI site of the cloning vector pBluescript (SK-) (pSK) to create the pSKHA plasmid.

To allow selection for the HA cassette in subsequent manipulations a tetracycline resistance gene (Tc^R) was cloned into the middle of the cassette. This Tc^R gene is intended to be removed before using the epitope, as it will interfere with expression. The Tc^R gene from the cloning vector pBR322 was amplified using the Tc5′ and Tc3′ primers that incorporate *Nco*I sites at either end of the gene. The PCR product was blunt cloned into the *Sma*I site of pSK to give the pSKTc^R plasmid. The Tc^R gene was then sub cloned into the *Nco*I site in the HA epitope sequence, in the same orientation as the epitope, to create the pSKHATc^R plasmid.

The HATc^R cassette was also sub cloned into the *Eco*RI site of the cloning vector pZErO. The cassette was sequenced with the M13 forward and reverse primers. The sequence of the 3x-HA epitope is shown in fig. 5.7.

5.2.7 Construction of the 6x-histidine 3x-HA double epitope cassette

As a useful tool for the one step tagging of chloroplast proteins with 6x-histidine residues and 3x-HA epitope sequences a double epitope cassette was constructed. This will allow the isolation of chloroplast proteins by IMAC and immunodetection using anti-HA antibodies. A summary of the cloning strategy is shown in fig 5.8. Six codons encoding for six histidine residues were added to the 3′ end of the HA cassette by a PCR strategy. The cassette was amplified using the HA(2.1) primer and a primer which is partially homologous to the 3′ end of the cassette, but which also contains a non-homologous tail containing 6 GTG codons encoding for six histidine residues (HAHis6). The PCR product was blunt cloned into the *Eco*RV site of the cloning vector pSK to give the pSKHA$Tc^R H_6$ plasmid. The cassette was sequenced with the M13 forward and M13 reverse primers to check that the

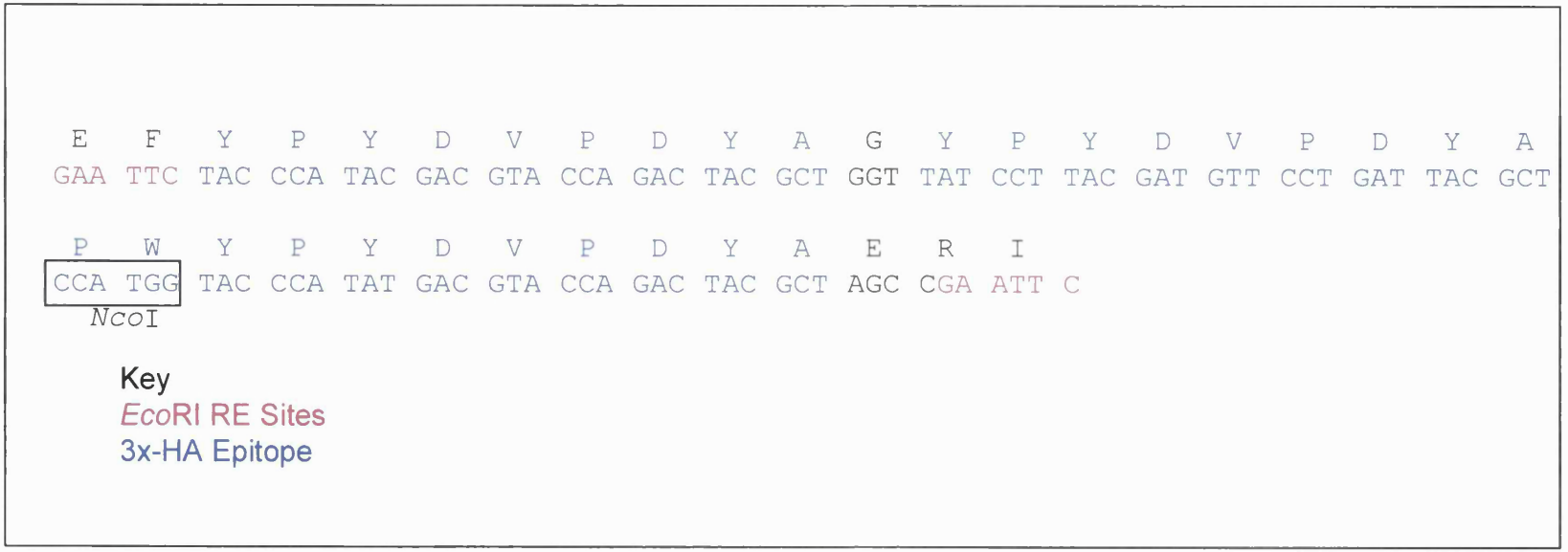


Fig. 5.7 Sequence of the 3x-HA epitope

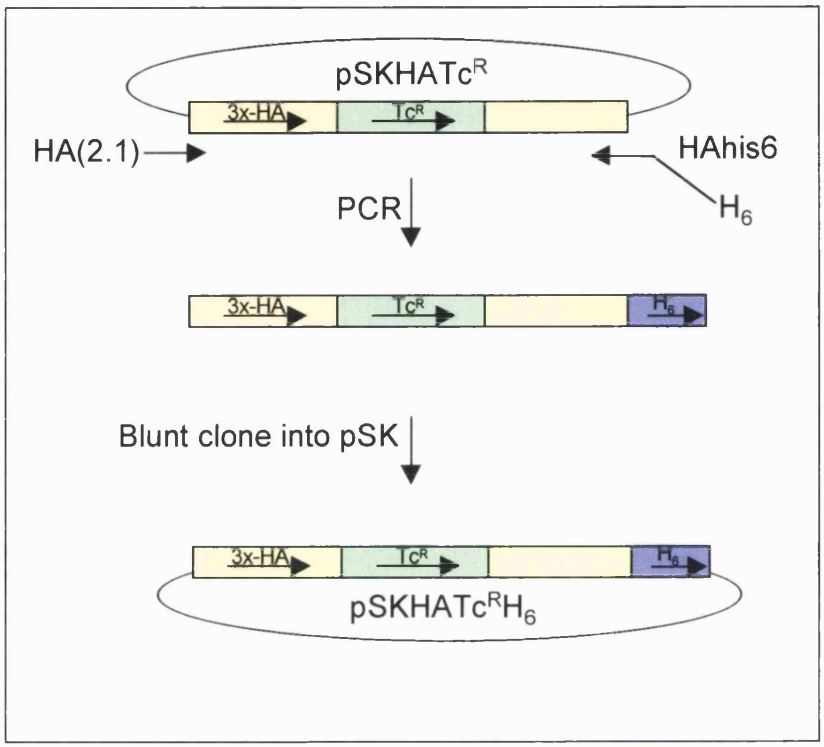


Fig. 5.8 Cloning strategy to tag the 3x-HATc^R cassette with 6x-histidine residues

3x-HA epitope and the 6x-histidine sequences were correct and in-frame (fig. 5.9).

5.2.8 Construction of the 3x-HA tagged β-galactosidase in pUEX2

In order to test that the new codon optimised 3x-HA epitope is immunoreactive the β-galactosidase protein was tagged with the epitope in the cloning vector pUEX2. This vector contains the β-galactosidase gene (lacZ) under the control of an inducible promoter. Furthermore a unique *Eco*RI site at the 3′ end of the gene allows the cloning of the 3x-HA cassette in frame with lacZ. The pSKHATc^R plasmid was digested with the *Eco*RI to drop out the 3x-HATc^R cassette. The 3x-HATc^R cassette was then cloned into the *Eco*RI site in the MCS of the expression vector pUEX2 (fig. 5.10) (Bressan & Stanley, 1987), downstream of, and in-frame with the lacZ gene encoding the β-galactosidase protein. This created the plasmid pUEX2HATc^R. The Tc^R gene was removed by digestion of pUEX2HATc^R with *Nco*I and the vector was re-ligated to produce the plasmid pUEX2HA. A fusion protein consisting of β-galactosidase and the 3x-HA epitope (β-gal-3x-HA) was encoded by this plasmid. The construct was sequenced with the pUEX2 primer to check that the 3x-HA tag was in-frame with the lacZ gene. The sequence generated is shown in fig. 5.11.

5.2.9 Western analysis of the β-gal-3x-HA protein

In order to verify that the new 3x-HA epitope cross reacts with commercially available monoclonal antibodies a western blot was carried out on the β-gal-3x-HA protein using anti-HA high affinity antibodies. Expression of the β-gal-3x-HA protein was induced by transferring a liquid culture of the *E. coli* strain DH5α containing the pUEX2HA plasmid from 30°C to 42°C for 2 hours. Expression of the β-gal protein was induced in the same way from DH5α containing the pUEX2 plasmid as a control. At 42°C the temperature sensitive

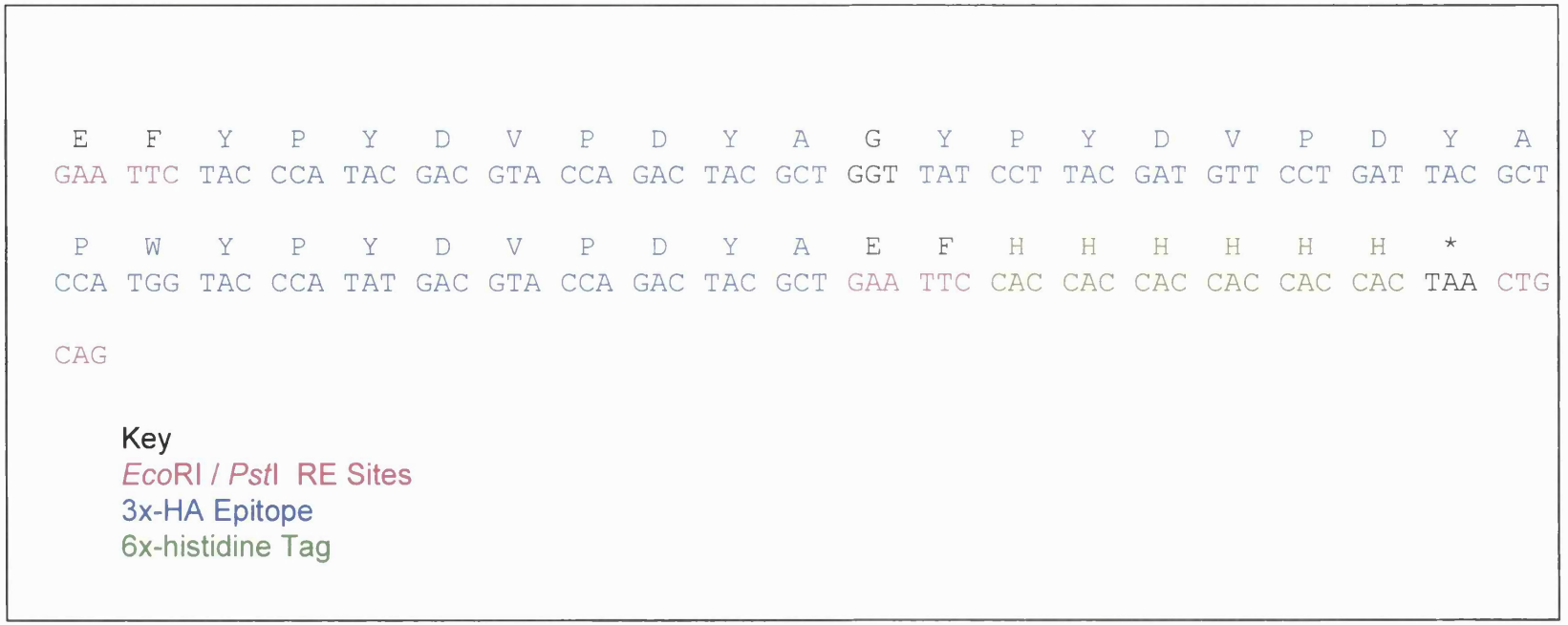


Fig. 5.9 Sequence of the 3x-HA 6x-histidine double epitope

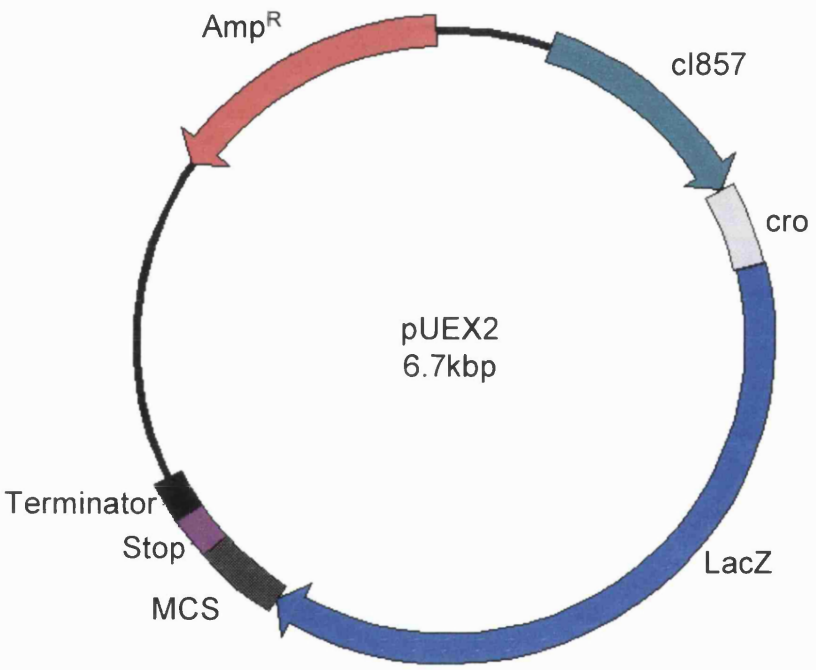


Fig. 5.10 Diagram of the pUEX2 plasmid showing the cI857 gene encoding a repressor protein, the cro promoter element (the target for the cI857 repressor), lacZ and Amp^R genes, the MCS and stop and terminator regions.

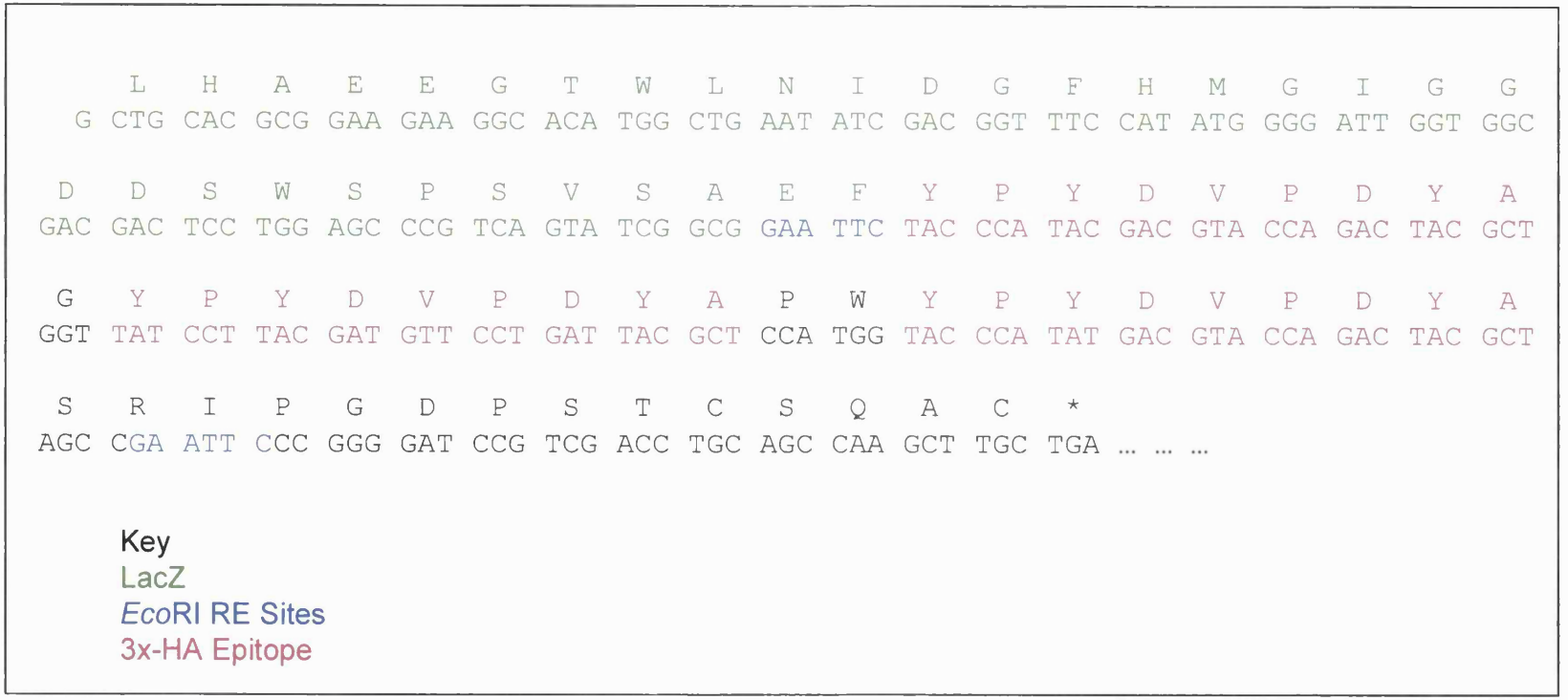
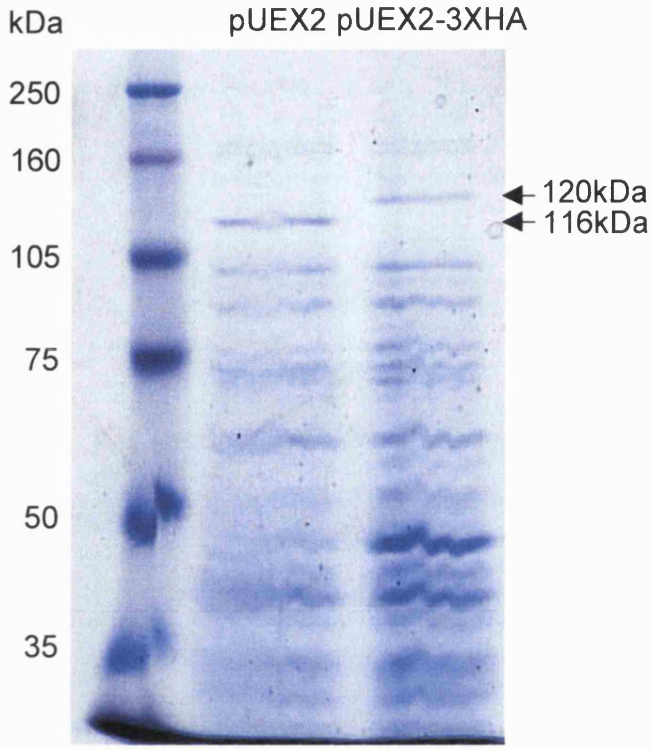


Fig. 5.11 Sequence of the pUEX2HA construct. The addition of the 3x-HA sequence increases the molecular mass of the β-galactosidase protein from 116kDa to 120kDa

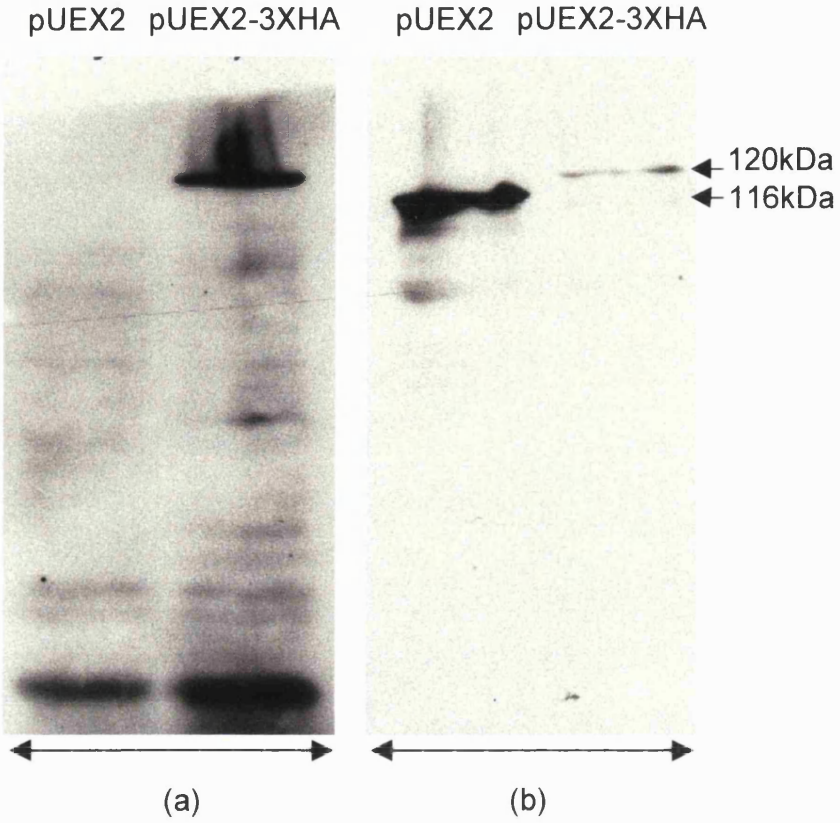
P_R repressor, encoded by the cI857 gene on the pUEX2HA plasmid is inactive, allowing expression of the β-gal-3x-HA gene (Stanley & Luzio, 1984). The protein was solubilised as described in 2.11.4 and run on a 7.5% SDS-PAGE gel. Fig 5.12 shows a coomassie blue stained SDS-polyacrylamide gel of the protein samples showing a band of 120kDa in the DH5α / pUEX2HA sample corresponding to the β-gal-3x-HA protein and a band of 116kDa in the DH5α / pUEX2 sample corresponding to the β-gal protein. The protein was transferred to nitrocellulose as described in 2.11.2 and probed with anti-HA high affinity primary antibodies and anti-rat IgG secondary antibodies. The optimal antibody concentrations and blocking / probing conditions had been determined by dot blots and are summarised in appendix II. A band of 120kDa is clearly seen in the DH5α / pUEX2HA sample corresponding to the β-gal-3x-HA protein, but no band is seen in the DH5α / pUEX2 sample as expected (fig. 5.13). The anti-HA antibodies specifically cross react with the codon optimised 3x-HA tag. Some non-specific cross reactivity of low molecular weight protein could be due to the secondary antibodies, or degradation products.

In order to verify that the protein detected by the anti-HA antibodies was the β-gal-3x-HA protein a western blot was carried out as before, using primary antibodies raised against β-galactosidase and anti-rabbit IgG secondary antibodies. Again, the optimal antibody concentrations and blocking / probing conditions had been determined by dot blots and are summarised in appendix II. In this case a band of 120kDa corresponding to the β-gal-3x-HA protein is seen in the DH5α / pUEX2HA sample and a band of 116kDa is seen in the DH5α / pUEX2 sample, corresponding to the β-gal protein (fig. 5.13). This confirms that the 120kDa band seen previously corresponds to the β-gal-3x-HA protein and also that the β-gal protein is present in the DH5α / pUEX2 sample, but that the anti-HA antibodies do not cross-react with it. It is unclear why the anti-β-gal antibodies react weakly with the β-gal-3x-HA protein compared to the β-gal protein, even though the proteins were equally



Total cell protein was solubilised from pUEX2 / DH5α and pUEX2-3XHA / DH5α and run on a 7.5% SDS-PAGE gel

Fig 5.12 SDS-PAGE of pEUX2 and pEUX2HA transformants



Total cell protein was solubilised from pUEX2 / DH5α and pUEX2-3XHA / DH5α and the filter was probed with (a) anti-HA 1° antibodies / anti-rat 2° antibodies and (b) anti-β-galactosidase 1° antibodies / anti-rabbit 2 ° antibodies.

Fig. 5.13 Western analysis of pEUX2 and pEUX2HA transformants

loaded. It demonstrates, however, the strong cross-reactivity of the anti-HA antibodies to the new 3x-HA epitope.

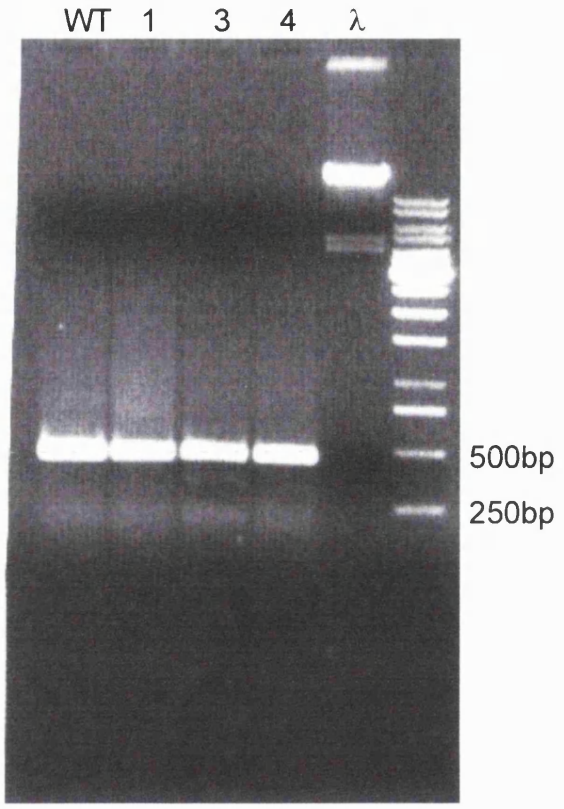
5.2.10 PCR analysis of putative psbH-HA transformants

In order to test that the 3x-HA epitope is immunoreactive in *C. reinhardtii* chloroplasts with minimum non-specific cross reactivity, the *psb*H gene encoding for the 9 kDa PSII phosphoprotein was chosen for tagging. A *psb*H-HA tagged construct was made by N. Ray (UCL) (fig 5.14) and used to transform a *psb*H-null mutant (*psb*H-null (opp) (O'Connor, 1998) by selection for phototrophic growth.

Total genomic DNA was extracted from three transformants and WT *C. reinhardtii* and primers were used to amplify the *psb*H gene. The resulting PCR product was digested with *Nco*I, as there is an *Nco*I site in the 3x-HA tag, but no *Nco*I site in the *psb*H gene. If the 3x-HA tag is present in the transformants two bands of 310bp and 130bp would be present, compared to a 440bp band in the WT. Fig. 5.14 shows the agarose gel of the digested PCR products. The bands in the transformant lanes are identical to the band in the WT lane suggesting that the *psb*H gene has integrated into the genome to rescue phototrophic growth, but the 3x-HA tag is absent. Fig. 5.15 shows the predicted recombination events. *psb*H is a small gene encoding a protein of just 9 kDa, it is possible that it will not tolerate such a large extension. The *psb*H product is an integral PSII protein and the epitope may affect PSII assembly or function.

5.3 Discussion

The chloroplast gene *rpo*C2 has been tagged with six codons coding for 6x-histidine residues. This was achieved by a simple PCR strategy involving the amplification of the 3′ end of the *rpo*C2 gene with two primers, one of which has a non-homologous tail consisting of the 6x-histidine tag nucleotide sequence. The *aad*A cassette was cloned downstream of the 6x-histidine



PCR products from genomic DNA extracted from WT and *psb*H-HA transformants amplified with H_2P and psbHi3′ primers, digested with *Nco*I. λ DNA was cut with *Nco*I as a control.

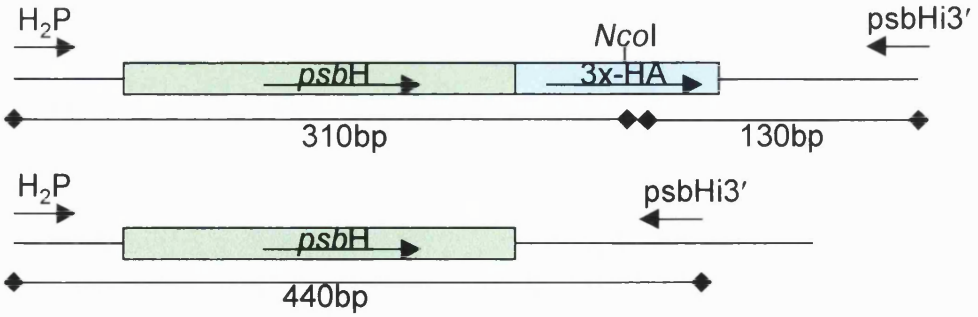


Fig. 5.14 PCR analysis of *psb*H-HA tagged transformants

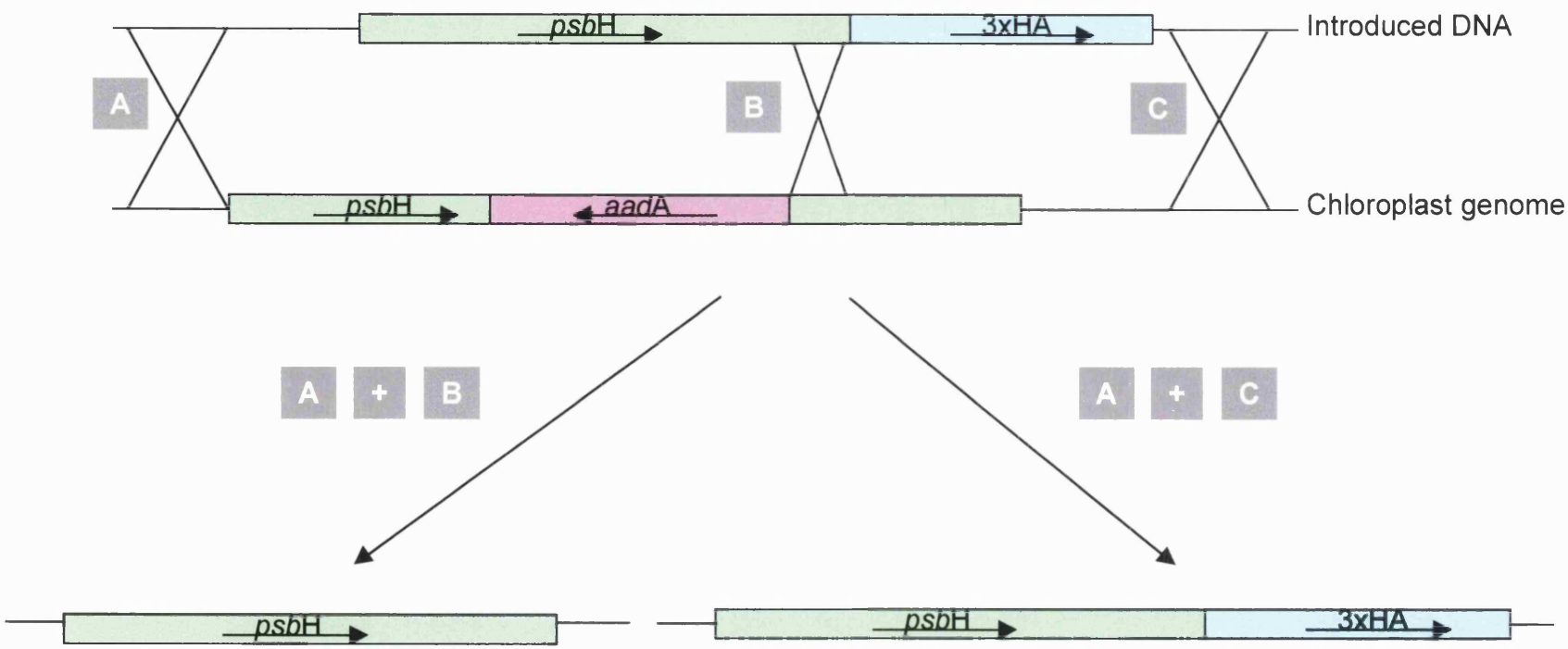


Fig. 5.15 Possible recombination events resulting from transformation of H-null (opp) with the psbH-HA construct.

tagged *rpo*C2 gene in order to allow for selection by spc^R. Transformants were obtained and have been shown, by Southern analysis, to contain both the *aad*A cassette and the 6x-histidine tag. The aim of this experiment was to use anti-his monoclonal antibodies in immunoblotting to detect and size the RpoC2 protein. The *rpo*C2 gene is unusually large and, when compared to its higher plant homologues, contains large insertion sequences that are in-frame with the rest of the gene. Other *C. reinhardtii* chloroplast genes also contain these insertions and it has been suggested that they may represent protein introns or inteins. In the case of the *clp*P gene a single amino-acid substitution at the putative intein splice junction allows the insertion to be spliced at the protein level in *E. coli* (Wang & Liu, 1997) providing strong evidence that this protein splicing does occur in *C. reinhardtii* chloroplasts. It was hoped that detection with anti-his monoclonal antibodies and sizing of the RpoC2 protein would allow the size *in vivo* to be compared to the size predicted from the DNA sequence (357.8kDa). Unfortunately, of the anti-his monoclonal antibodies tested, none showed any immunoreactivity to the $RpoC2H_6$ protein or to other 6x-histidine tagged proteins.

As an alternative method to detect and size the RpoC2 protein the affinity of the 6x-histidine tag for Ni^{2+} was used in an attempt to isolate the protein from a *C. reinhardtii* membrane preparation. Preliminary experiments could not detect a protein of 340 kDa, but detected a 105kDa band by SDS-PAGE. It is possible that this represents a spliced version of the RpoC2 protein. Further experiments are necessary to conclusively determine this. It would be useful to carry out immunoblotting on the purified protein using anti-His monoclonal antibodies. In light of the lack of immunoreactivity of the anti-His antibodies it may be preferable to tag the *rpo*C2 gene with the $3x\text{-}HAH_6$ double epitope. The protein could be purified as before by IMAC, then detected using anti-HA monoclonal antibodies.

A new 3x-HA tag that is codon-optimised for use in *C. reinhardtii* chloroplasts has been successfully synthesised *de novo* by overlap PCR. An advantage of

this method is that a double stranded DNA molecule can be synthesised without the need for a template. The double stranded nature of the product was tested by RE digestion. REs will only digest double stranded DNA molecules. The RE site is situated on a region of the HA-3′ primer that does not overlap with the HA-5′ primer, so successful digestion will only result if the DNA polymerase has extended from the overlapping primers to create a double stranded molecule.

The cloning of a removable antibiotic resistance marker (Tc^R) into the middle of the tag is a useful feature in that it will allow the tag to be easily inserted into plasmids containing cloned chloroplast genes as it increases the size of the cassette, making it portable and allows selection. Most commonly used plasmids carry genes conferring resistance to ampicillin (Amp), so selection on Tc containing media provides an instant screen for recombinant plasmids containing the 3x-HATc^R insert. Simple digestion with *Nco*I and re-ligation allows the Tc^R gene to be removed and a fully functional 3x-HA tag to be obtained. The 3x-HATc^R cassette has also been cloned into the cloning vector pZErO, which lacks the Amp^R gene. If this plasmid is used to obtain the 3x-HATc^R cassette by PCR or RE digestion it is not necessary to purify the PCR product or RE fragment away from the "parent" plasmid before ligation into the cloned chloroplast gene. Selection on Amp + Tc will screen for the plasmid containing the cloned chloroplast gene and the 3x-HATc^R cassette. It is hoped that these features will make manipulation of chloroplast genes with the 3x-HA tag rapid and simple. Additional features that would enhance the tag and make it easier to use would be the engineering of a multiple cloning site (MCS) at the 5′ end. A series of constructs with the 3x-HA tag in all 3 reading frames would be useful and, along with the MCS allow greater flexibility of cloning.

The 3x-HAH_6 double epitope would also benefit from these features. In addition a protease cleavage site could be engineered at the beginning of the

6x-histidine tag. This would allow the 6x-histidine tag to be removed after purification of the protein. The 3x-HA tag would remain to allow immunoanalysis of the purified protein.

The HA epitope was chosen from a number of possible epitopes to develop for use in *C. reinhardtii* chloroplasts as it has been used to tag *C. reinhardtii* nuclear genes. Immunological analysis of the HA tagged proteins has shown that *C. reinhardtii* proteins do not non-specifically cross react with anti-HA monoclonal antibodies (Boudreau *et al.*, 2000), (Kozminski *et al.*, 1993). The new codon optimised 3x-HA tag has been shown, by tagging of β-galactosidase in *E. coli*, to cross react with the anti-HA monoclonal antibodies. At optimal antibody dilutions non-specific cross-reactivity of the anti-HA monoclonal antibodies to *E. coli* proteins is low, allowing clear detection of the 3x-HA tagged β-galactosidase. For these reasons it is hoped that this epitope will be used to tag *C. reinhardtii* chloroplast proteins such as orf1995 an essential protein of unknown function (Boudreau *et al.*, 1997). The tag could also be used to tag chloroplast genes in higher plants or other algae. Possible candidates are the large number of orfs of unknown function in red algae such as *Porphyra* (1.2.3).

It was hoped that tagging of the PSII gene, *psb*H would allow a clear demonstration of the utility of this epitope in the tagging of *C. reinhardtii* chloroplast genes. Unfortunately, PCR analysis of the transformants showed that the 3x-HA tag had not incorporated into the chloroplast genome at the desired location. The *psb*H gene is small, encoding a protein of just 9 kDa. In retrospect it is highly likely that this protein would be unable to tolerate an extension of 4kDa. Also, the *psb*H-HA tagged transformants were selected by phototrophic growth. It is possible that these conditions were too stringent as they rely on fully functional photosystems and encourage the integration of the *psb*H gene without the 3x-HA tag. The exact function of PSII-H within PSII is unclear, and the 3x-HA tag could well interfere with this function. An

alternative strategy could involve cloning the *aad*A cassette downstream of the *psb*H-HA gene and transforming WT *C. reinhardtii*, selecting for spc^R. Although the transformants may not be photosynthetic, the *psb*H-HA gene could still be expressed and the PSII-H-HA protein produced, but the protein would be too large to assemble into functional PSII. Nonetheless, The PSII-H-HA protein could still be detected by immunoanalysis.

Alternatively a different chloroplast gene could be chosen for tagging with the 3x-HA epitope. The putative nucleoid binding protein ORF2971 (A. T. Watson, unpublished data) is a very large gene encoding a protein of 2971 amino-acids. This large protein is more likely to tolerate an addition of 4kDa, and a transformation strategy based on the *aad*A cassette and spc^R would encourage integration of the 3x-HA tag. Immunological analysis of the protein could provide further evidence for its function in nucleoid binding.

Chapter 6 – Discussion

6.1 Addressing the aim of this thesis

The title of this thesis is "The transcriptional apparatus of *Chlamydomonas* chloroplasts" and the work described in the three results chapters involve experiments designed to further the understanding of the genes and their corresponding proteins involved in the transcription of the chloroplast genome in this green alga. Despite a large amount of work on the transcriptional apparatus of higher plant chloroplasts in recent years many questions remain over the situation in green algae. The model organism *C. reinhardtii* is an ideal system in which to answer these questions due to the ease of chloroplast transformation in this species, the availability of selectable markers for chloroplast reverse-genetics, the identification of chloroplast genes involved in chloroplast transcription and the large scale EST projects aimed at characterising the *C. reinhardtii* nuclear genome.

In chapter three the aim was to characterise the PEP of *C. reinhardtii* in order to gain information about the presence of a NEP. In higher plants the chloroplast genome is transcribed by two polymerases, PEP and NEP (Hess & Borner, 1999). Previous studies indicate that *C. reinhardtii* lacks a NEP (Rochaix, 1997), and this is consistent with the theory on the evolutionary origin of NEP (Gray & Lang, 1998). Reverse-genetics using the chloroplast gene *rpo*C2, which encodes the β″ subunit of PEP allows insights into the nature of chloroplast transcription in *C. reinhardtii.* The use of chloroplast RNA polymerase inhibitors provides an alternative and complementary approach to answering these questions.

In chapter four the aim was to clone and sequence *C. reinhardtii* nuclear genes which are involved in chloroplast transcription. A large number of nuclear genes involved in chloroplast transcription have been cloned and sequenced from the nuclear genomes of higher plants and red algae (Tozawa *et al.*, 1998), (Isono *et al.*, 1997), but to date none have been cloned in *C. reinhardtii.* Genes encoding PEP σ factors have been identified in both

higher plant and red algal species and it seems likely that the nuclear genome of *C. reinhardtii* contains genes encoding these proteins. The *rpo*A gene, encoding the essential α subunit of PEP, is normally chloroplast encoded, but is absent from the almost completely sequenced chloroplast genome of *C. reinhardtii*. It is possible that, as in the case of the malaria parasite *Plasmodium falciparium*, the *rpo*A gene has been transferred to the nucleus (Sato *et al.*, 2000). Being the smallest PEP gene, *rpo*A is the most likely candidate for this event, giving the nucleus extra control over chloroplast transcription. RACE and searches of the EST database were employed to identify these genes in the nuclear genome of *C. reinhardtii*. Finally efforts will be made to identify a NEP gene in the nuclear genome of *C. reinhardtii* by PCR and searches of the EST database.

In chapter five the aim was to analyse the RpoC2 protein in more detail by epitope tagging the protein with 6x-histidine residues which can be exploited in IMAC. This would allow further characterisation of PEP and the RpoC2 protein. As an alternative approach a 3x-HA epitope could be used to tag the RpoC2 protein but would need to be codon optimised for chloroplast expression.

6.2 Discussion of chapter 3 results and future work

Following the re-sequencing of the *C. reinhardtii rpo*B and *rpo*C2 genes encoding the PEP β and β″ subunits it seemed necessary to make new knockout mutants now that the genes had been properly defined. Previous work had indicated that disruption of PEP genes in *C. reinhardtii* yielded heteroplastic transformants, suggesting that these genes are essential to cell viability (Rochaix, 1997). This is in direct contrast to the situation in tobacco where homoplastic transformants can be obtained and the genes are dispensable (Allison *et al.*, 1996) and in non-photosynthetic parasitic plants where the *rpo* genes are absent from the plastid genome but chloroplast transcription is still detectable (dePamphilis & Palmer, 1990). Gene

knockouts using the *C. reinhardtii aad*A cassette to disrupt the correctly sequenced *rpo*C2 gene confirmed this and these findings were reinforced by the observation that the selective pressure for functional *rpo*C2 is so strong that in the absence of spectinomycin the *aad*A cassette is rapidly lost.

The experiments performed differed fundamentally, however. The genes were knocked out using *aad*A cassettes driven from different promoters. The tobacco knockouts were made using an *aad*A cassette driven from the *rrn*S gene promoter, a housekeeping gene (Svab & Maliga, 1993). The *C. reinhardtii* knockouts were made using an *aad*A cassette driven from the *atp*A gene promoter (a photosynthetic gene). This is significant because in higher plants PEP is known to specifically transcribe photosynthetic genes. Disrupting a PEP gene with an antibiotic resistance cassette driven from a photosynthetic gene promoter may be impossible, as PEP is needed to transcribe the cassette, and provide drug resistance. The gene may not be necessary for cell viability *per se*. In order to overcome this the promoter of the existing *aad*A cassette was replaced with that of the promoter from the *C. reinhardtii rrn*S gene. This new cassette was used to disrupt *rpo*C2 as before. Gene disruption transformants could not be isolated possibly due to problems with the new promoter. Although the new promoter contained –10 and –35 sequences which resemble prokaryotic promoters and a transcription start site and the transcription start site had been mapped by S1 nuclease digestion, it is possible the transcript had been processed and the true transcription start site is further upstream. Also the new *rrn*S promoter is a chimeric structure. Elements upstream of the start site needed for efficient translation may be absent as *rrn*S encodes for ribosomal RNA rather than protein. It is quite possible that the promoter of the new *aad*A cassette is non-functional in *C. reinhardtii*. Useful experiments to test this would be to knock out a gene which can readily be disrupted such as *psb*H (O'Connor *et al.*, 1998) or inserting the cassette in a non-coding region (such as downstream of *psb*H), then transforming a Δ*psb*H mutant. Transformants can be selected under phototrophic conditions and northern analysis can be used to

determine whether the marker is expressed. If the cassette is non-functional a new promoter, possibly from a different housekeeping gene, could be used to repeat the experiment and further clarify the essential nature of PEP genes in *C. reinhardtii.*

The alternative approach of knocking the gene out using a non-expressed cassette lends weight to the theory that PEP genes are essential to *C. reinhardtii* cell viability. Non-essential genes can routinely be knocked out by a co-transformation approach using a non-expressed piece of "stuffer" DNA to disrupt the gene and a variant of the *rrn*S gene which confers resistance to spectinomycin and streptomycin (Kindle *et al.*, 1991), (Newman *et al.*, 1991). If *rpo*C2 is non-essential then this gene could also be successfully disrupted by this method. However, gene disruption transformants could not be isolated suggesting that the selective pressure for functional *rpo*C2 is so strong that a disruption will not be tolerated, except if there is an equally strong selective pressure for that disruption. The advent of a second marker (*aph*A-6) for chloroplast transformation conferring kanamycin resistance (Bateman & Purton, 2000) now allows this to be tested. A co-transformation is performed using *aph*A-6 and the spectinomycin resistant 16S rRNA gene in which *rpo*C2 is disrupted by *aph*A-6. Primary transformants are selected on both drugs. Spectinomycin only is then used in an attempt to drive the knockout to homoplasmicity.

Given the recent advances in chloroplast transformation in other green algal species such as *Euglena gracilis* and red algae such as *Porphyridium* (Stevens & Purton, 1997) interesting future experiments would be to knock out the *rpo* genes in other algal species. This would identify a major difference between the transcriptional apparatus of higher plant and algal species. In *E. gracilis* the *rpo* genes have been cloned and sequenced, and an *aad*A cassette has recently been developed as a selectable marker in this alga (R. Hallick, pers. comm.).

The use of PEP inhibitors also provides evidence that PEP is essential to *C. reinhardtii* cell viability. Of the RNA polymerase inhibitors tested on WT and cell wall deficient *C. reinhardtii* only rifampicin and actinomycin D had an inhibitory effect. Actinomycin D is a general RNA polymerase inhibitor and was used as a control. Rifampicin, however has an inhibitory effect on both heterotrophic and phototrophic growth. The minimum inhibitory concentration of rifampicin is the same, 100μg / ml, in both heterotrophic and phototrophic conditions. If the situation in *C. reinhardtii* were the same as in higher plants then it would be predicted that rifampicin would prevent phototrophic, but not heterotrophic growth. If rifampicin prevents heterotrophic growth, then it follows that PEP is essential.

The transcription assay provides the only evidence for a second RNA polymerase activity in *C. reinhardtii* chloroplasts. In the presence of rifampicin transcription of the *rrn*S gene can be detected. This suggests that a second rifampicin insensitive RNA polymerase is present (possibly a NEP). Recent work carried out in mustard however, suggests that the PEP can exist in rifampicin sensitive and rifampicin insensitive forms, which are interchangeable by phosphorylation of the complex (Pfannschmidt *et al.*, 2000). In light of the previous experiments, which strongly suggest that PEP is essential, it seems likely that the rifampicin insensitive RNA polymerase activity, which transcribes the *rrn*S gene, is the result of a phosphorylated form of PEP. If an inhibitory effect can be determined using other, more specific PEP inhibitors such as tagetitoxin, by increasing the concentration tested, the transcription assay could be repeated using this inhibitor to detect a second RNA polymerase activity.

6.3 Discussion of chapter 4 results and future work

Following searches of the EST database, two cDNA clones were identified which showed significant similarity to PEP σ factors. Sequencing revealed that these cDNA clones corresponded to the partial cDNAs from the same

gene, *rpo*D. Southern analysis determined that the *rpo*D gene is present as a single copy in the nuclear genome of *C. reinhardtii*, and does not identify a gene family of σ factors. Northern analysis of total RNA extracted from *C. reinhardtii* cells grown in mixotrophic, heterotrophic and phototrophic conditions revealed a transcript of ~2.9kbp which is present in RNA extracted from cells grown in heterotrophic and phototrophic conditions, but absent from cells grown in mixotrophic conditions. The transcript is most abundant in cells grown in phototrophic conditions. This is to be expected, as PEP has a role in phototrophic growth, but the absence of the transcript in cells grown in mixotrophic conditions is difficult to explain. The presence of acetate in growth media is known to reduce CO_2 fixation and O_2 evolution, but this does not explain the presence of the transcript in cells grown in heterotrophic conditions. More experiments need to be carried out to investigate this result further and to confirm the size of the transcript. For example RT-PCR could be used to obtain a complete cDNA, containing the 5′ coding sequence and 5′ UTR and compare the size to that predicted from the northern blot. RT-PCR can be carried out from mRNA extracted from cells grown in mixotrophic, heterotrophic and phototrophic conditions and the results compared to those of the northern blot.

Screening of a genomic library identified two cosmids containing the *rpo*D genomic sequence. A 4.6kbp *Hind*III fragment containing part of the *rpo*D gene was subcloned and sequenced. *rpo*D is a typical *C. reinhardtii* gene in terms of GC content, stop codon, polyadenylation signal, 3′ UTR, intron size and consensi at the 5′ and 3′ splice sites (Silflow, 1998). Further work needs to be carried out to clone and sequence the remaining 5′ coding region and 5′ UTR in order to obtain a complete genomic sequence. When compared to a complete cDNA sequence the deduced protein sequence and the presence of introns can be conclusively determined.

From the known coding sequence several predictions can be made regarding the protein although a complete cDNA and genomic sequence are needed to obtain more information about the *N*-terminus of the protein. For example the protein is at least 72,593 Da (73 kDa), and consists of at least 678 amino acids. No known chloroplast targeting peptide has been predicted by either the ChloroP or the SignalP programes, although alignments with cyanobacterial σ factors reveal an *N*-terminal extension. Although *C. reinhardtii* transit peptides differ significantly from those of higher plants (Franzen, 1995) the lack of an obvious transit peptide is most likely the result of the lack of sequence data from the 5′ end of the *rpo*D gene, resulting in a protein which is truncated at the *N*-terminus. Again, a complete cDNA and genomic sequence should resolve this.

The sequencing of the *rpo*D gene reveals why the σ factors RACE failed. The primers differed in nucleotide sequence at vital positions to the *rpo*D gene and it appears from subsequent northern analysis that the gene is not expressed in cells grown in TAP in the light. This is the most likely reason for the failure of the *rpo*A RACE. *rpo*A encodes an essential structural subunit of PEP (the α subunit), and if it is lacking from the chloroplast genome it must be present in the nucleus. There is a precedent for this gene translocation event in that the *rpo*A gene has recently been identified in the nucleus of the malaria parasite *Plasmodium falciparum* (Sato *et al.*, 2000). An alternative method of identifying the *rpo*A gene in the *C. reinhardtii* nuclear genome could be to carry out a heterologous Southern using an *rpo*A gene from a bacterium with a GC rich genome to probe the blot. Regular searches of the EST database have failed to identify a putative *rpo*A gene, but given time and the complete sequencing of the genome an *rpo*A gene may be identified. The failure to identify a NEP, however, is more complex to explain. From the results presented in chapter 3 it seems highly likely that *C. reinhardtii* has no NEP. However, the failure of the PCR could also be attributed to the reasons

cited for the failure of the *rpo*A and σ factors RACE. Again a completely sequenced nuclear genome is necessary to resolve this question.

6.4 Discussion of chapter 5 results and future work

The *rpo*C2 gene is unusually large and the deduced amino acid sequence shows blocks of homology to the *rpo*C2 gene of higher plants, separated by extra regions which are in-frame with the rest of the gene. As such this gene is a candidate for protein splicing. In order to investigate this the gene was tagged with six codons coding for six histidine residues at the 3′ end of the gene. It was hoped that antibodies raised against the 6x-histidine tag could be used in western analysis to size the RpoC2 protein. The commercially available anti-his antibodies are notoriously unreactive, and no cross reactivity could be detected with either the 6x-histidine tagged RpoC2, or other 6x-histidine tagged proteins, including ones which had been purified by IMAC. As an alternative cell extracts from the 6x-histidine tagged transformants were subjected to IMAC in order to purify the RpoC2 protein as a single polypeptide or as part of the PEP complex. SDS-PAGE revealed that, despite a large amount of non specific binding of *C. reinhardtii* proteins to the Ni^{2+}-NTA resin, a single band of ~105 kDa. No band corresponding to the predicted size of RpoC2 (357.8 kDa) could be detected. It is possible that the 105 kDa band corresponds to post translationally spliced RpoC2, but further work needs to be carried out in order to confirm this. If, for example, the band specifically cross reacted with the anti-his antibodies, then it could be concluded to correspond to the RpoC2 protein.

The codon optimised 3x-HA tag was developed as an alternative to the 6x-histidine tag to tag *rpo*C2 and detect the protein by western analysis, again to size the protein. This tag was designed as a mobile element, with the insertion of the Tc^R gene to increase the overall size of the cassette and allow direct selection. The tag has been shown to be highly immunoreactive with the commercially available anti-HA antibodies, as these cross react with

3x-HA tagged β-galactosidase, when over expressed from the pUEX2 vector. It was hoped that the highly expressed photosynthetic gene *psb*H could be tagged with the new 3x-HA tag in order to test the reactivity of the tag in *C. reinhardtii*, and to provide a control for western analysis using the poorly expressed *rpo*C2 gene. The size of the *psb*H gene, however, proved too small to tolerate the addition of this 100bp tag. A larger gene, such as the putative nucleoid binding protein, *orf*2971, would be a candidate to tag in the future. The anti-HA antibodies could also be used in order to elucidate more regarding the function of *orf*2971. For example, to visualise, by immunogold labelling, the binding of *orf*2971 to nucleoids.

The 6x-histidine 3x-HA double epitope was developed as a tool to allow purification of a protein by IMAC, followed by immunoanalysis using anti-HA antibodies. An important experiment would be to tag the *C*-terminus of the RpoC2 protein with this double epitope in order to purify the protein by IMAC, as before, then detect the protein by western analysis. As the *rpo*C2 gene is poorly expressed it seems likely that the RpoC2 protein will be of too low abundance to be detected from crude cell lysate, so an initial purification would be necessary.

An alternative approach to investigating protein splicing of the *rpo*C2 gene would be to size the transcript and compare that to the size of the gene. The fact that the *rpo*C2 gene extends as a single open reading frame in the AT rich chloroplast genome, strongly suggests that there are no introns present, but the possibility cannot be ruled out. RT-PCR to isolate a complete cDNA sequence would answer this question.

6.5 Concluding remarks

In this thesis a combination of both molecular-genetic and biochemical studies have been used in an attempt to shed new light on an elaborate

transcriptional system that has evolved in the plastid in response to the need of the eukaryotic cell to co-ordinate and regulate plastid biogenesis.

Appendices

Appendix I – Oligonucleotide primers

DNA oligonucleotide primers used for PCR and DNA sequencing are listed. Sequences are shown in the 5'→3' direction.

I.i Primers used in chapter 3

Primer Name	Primer Sequence
rrnSP$_R$5′	ctagaattcaaatatcggcag
rrnSP$_R$3′(2)	attcatgaacataattttagctgtatggggtattagc
Tc5′	tcccatggttgacagcttatcatcg
Tc3′	ccaatccatggagtggtgaatcc

I.ii Primers used in chapter 4

Primer Name	Primer Sequence
oligodT-58	aaggatccgtcgacatcgataatacgactcactataagggattttttttttttttttt
RACE 1	aaggatccgtcgacatcgataat
RACE 2	gataatacgactcactataaggga
Sigma(F)	tactggtggatycgbcagg
SigmaN	cgbgagcgbgtscgbcag
rpoA(F)	ctsgagatyctsctsaacaa
rpoA.N	gtsacsgcbgcbgacmt
NEPGN(F)	ggbttytactacccbcacaa
NEPVT(F)	gtbaagcagacbgtbatgac
NEPVT(R)	gtcatvacvgtctgcttvac
NEPCA(F)	tgcaacggbctscagcactacgc
NEPCA(R)	gcgtagtgctgsagvccgttgct
D392(F)	gcagtttgccgcgacttcc
D392(R)	gcgccagctgcgtctgc
D392(Int)	gcagacgcagctggcgc
D392(Int)2	acgctacctcatgcaatgacta
D392(Int)3	caagaacctgtccgccaagct
D392(Int)4	ggatgcgctcgcgtgttacg
D390(F)	cggctgagtctgaggacg

D390(R)	cgcgcagccatctcgcgc
D390(Int)	gcgcgagatggctgcgcg
D390(Int)2	agagaggcttagagcacctga
D390(Int)3	acgctctcttagtgctgaca
D1	accaacgagccggtgaaccag
D2	cgacgaaggctgctgctgc
D3	gtgcagcgtcttctcgcact
D4	tcagcggtgcgttgctcgca
D5	tgcacggtaagcacagaggc
D6	gtgacgacttgcgcgacacct
D7	atagctgcttttgtctgcagcc
D8	cgctacgcatgccctgcca
D9	caggtaaggagctgatgtgc
D10	atcgtgagtgctacggcgga
D11	catattccggtcccttgtgctt
D12	agagcataaggcatagctggct
D13	ttcaccgtgtagacagtgagca
D14	ggtagctggagggggcagg
D15	atgaggttgttcggcctcaga
D16	ttcaggattggcatagtttc
D17	cagcatgtagcgtgcaggctt
D18	tgcgcaagctgcgtgtgaag
D19	gatccaccagtgcgctgtac
D20	tccgacgtgcagcacctgta
D21	tgctgcacgtcgaaagctgct
T3	aattaaccctcactaaaggg
T7	gtaatacgactcactatagggc

I.iii Primers used in chapter 5

Primer Name	Primer Sequence
rpoC-PmI	gtattagggattacacgtgc
rpoC-his	gaaggccttagtggtggtggtggtgactagataaaaaac
HA-5′	ctacccatacgacgtaccagactacgctggttatccttacgatgttcctgattacgctcc
HA-3′	cggctagcgtagtctggtacgtcatatgggtaccatggagcgtaatcaggaacatcg
HA(2.1)	gtttgggaattctacccatacgacg

HA(2.2)	tgggttgaattcggctagcgtagtc
Tc5′	tcccatggttgacagcttatcatcg
Tc3′	ccaatccatggagtggtgaatcc
M13 reverse	ggaaacagctatgaccatg
M13 forward	gtaaaacgacggccagt
HAHis6	ctgcagttagtggtggtggtggtggaattcagcgtagtctggtacg
pUEX2	tccagttcaacatcagccgc
H_2P	tcaattatggcaacagg
psbHi3′	cccgggatccaagaaaagtgagctattaacg

Appendix II – Western blotting conditions

Optimal antibody dilutions, blocking and probing conditions for immunodetection as determined by dot blots are listed along with the sources of antibodies.

Antibody	Optimal Dilution	Optimal Blocking Conditions	Optimal Probing Conditions	Source
anti-His (C-term)	-	-	-	Invitrogen
tetra His	-	-	-	Qiagen
anti-HA	1:10,000	1% BSA in TBS	1% BSA in TBS	Boehringer Mannheim
anti-rat	1:10,000	1% BSA in TBS	1% BSA in TBS	Sigma
anti-β-gal	1:10,000	5% milk in TBS-T	1% milk in TBS-T	Prof. J. Gray
anti-rabbit	1:10,000	5% milk in TBS-T	1% milk in TBS-T	Amersham

References

Allison, L. A. (2000). The role of sigma factors in plastid transcription. *Biochimie* **82**, 537-548.

Allison, L. A., Simon, L. D. & Maliga, P. (1996). Deletion of *rpo*B reveals a second distinct transcription system in plastids of higher plants. *EMBO J.* **15**, 2802-2809.

Apel, K. & Bogorad, L. (1976). Light-induced increase in the activity of maize plastid DNA-dependent RNA polymerase. *Eur. J. Biochem.* **67**, 615-620.

Armstrong, J. J., Surzycki, S. J., Moll, B. & Levine, R. P. (1971). Genetic transcription and translation specifying chloroplast components in *Chlamydomonas reinhardtii. Biochemistry* **10**, 692-701.

Arnold, F. H. & Haymore, B. L. (1991). Engineered metal-binding proteins: Purification to protein folding. *Science* **252**, 1796-1797.

Aro, E.-M., Virgin, I. & Andersson, B. (1993). Photoinhibition of Photosystem II. Inactivation, protein damage and turnover. *Biochim. Biophys. Acta* **1143**, 113-134.

Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. & Tabata, S. (1999). A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii.* Generation of 3433 non-redundant expressed sequence tags. *DNA Res.* **6**, 369-373.

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**, 181-189.

Bateman, J. M. & Purton, S. (2000). Tools for chloroplast transformation in *Chlamydomonas*: Expression vectors and a new dominant selectable marker. *Mol. Gen. Genet.* **263**, 404-410.

Baumgartner, B. J., Rapp, J. C. & Mullet, J. E. (1993). Plastid genes encoding the transcription / translation apparatus are differentially transcribed early in barley (*Hordeum vulgare*) chloroplast development. *Plant Physiol.* **101**, 781-791.

Bertioli, D. (1997). Rapid Amplification of cDNA Ends. In *PCR Cloning Protocols*, pp. 233-238. Edited by B. A. White. New Jersey: Humana Press Inc.

Bottomley, W. (1970). Deoxyribonucleic acid-dependent ribonucleic acid polymerase activity of nuclei and plastids from etiolated peas and their response to red and far red light *in vivo*. *Plant Physiol.* **45**, 608-611.

Bottomley, W. (1970). Some effects of Triton X-100 on pea etioplasts. *Plant Physiol.* **46**, 437-441.

Bottomley, W., Smith, H. J. & Bogorad, L. (1971). RNA polymerases of maize: Partial purification and properties of the chloroplast enzyme. *Proc. Natl. Acad. Sci. USA* **68**, 2412-2416.

Bottomley, W., Spencer, D., Wheeler, A. M. & Whitfield, P. R. (1971). The effect of a range of RNA polymerase inhibitors on RNA synthesis in higher plant chloroplasts and nuclei. *Arch. Biochem. Biophys.* **143**, 269-275.

Boudreau, E., Nickelsen, J., Lemaire, S. D., Ossenbuhl, F. & Rochaix, J.-D. (2000). The *Nac2* gene of *Chlamydomonas reinhardtii* encodes a

chloroplast TPR-like protein involved in *psb*D mRNA stability. *EMBO J.* **19**, 3366-3376.

Boudreau, E., Turmel, M., Goldschmidt-Clermont, M., Rochaix, J.-D., Sivan, S., Michaels, A. & Leu, S. (1997). A large open reading frame (*orf1995*) in the chloroplast DNA of *Chlamydomonas reinhardtii* encodes an essential protein. *Mol. Gen. Genet.* **253**, 649-653.

Boudreau, E., Otis, C. & Turmel, M. (1994). Conserved gene clusters in the highly rearranged chloroplast genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii. Plant Mol. Biol.* **24**, 585-602.

Boyer, A. S. & Hallick, R. B. (1998). Purification and characterisation of a soluble DNA-dependent chloroplast RNA polymerase from *Pisum sativum. Plant Sci.* **137**, 13-32.

Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., Johnson, A. M., Jones, A. R., Randolph-Anderson, B. L., Robertson, D., Klein, T. M., Shark, K. B. & Sanford, J. C. (1988). Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* **240**, 1534-1538.

Bressan, G. M. & Stanley, K. K. (1987). pUEX, a bacterial expression vector related to pEX with universal host specificity. *Nucleic Acids Res.* **15,** 10056.

Briat, J.-F., Laulhere, J.-P. & Mache, R. (1979). Transcription activity of a DNA-Protein complex isolated from spinach plastids. *Eur. J. Biochem.* **98**, 285-292.

Briat, J.-F. & Mache, R. (1980). Properties and characterization of a spinach chloroplast RNA polymerase isolated from a transcriptionally active DNA-protein complex. *Eur. J. Biochem.* **111**, 503-509.

Chang, C.-C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S. & Stern, D. B. (1999). Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. *Plant Cell* **11**, 911-926.

Christopher, D. A., Kim, M. & Mullet, J. E. (1992). A novel light-regulated promoter is conserved in cereal and dicot chloroplasts. *The Plant Cell* **4**, 785-798.

Cooper, A. A. & Stevens, T. H. (1993). Protein splicing: Excision of intervening sequences at the protein level. *BioEssays* **15**, 667-674.

Cullin, C. & Minvielle-Sebastia, L. (1994). Multipurpose vectors designed for the fast generation of N- or C- terminal epitope-tagged proteins. *Yeast* **10**, 105-112.

Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J. & Sedgwick, S. G. (1992). Protein splicing in the maturation of *M. tuberculosis* RecA protein: A mechanism for tolerating a novel class of intervening sequence. *Cell* **71**, 201-210.

Davis, E. O., Thangaraj, H. S., Brooks, P. & Colston, M. J. (1994). Evidence of selection for protein introns in the RecAs of pathogenic mycobacteria. *EMBO J.* **13**, 699-703.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist* **154 (supp.)**, S164-S177.

dePamphilis, C. W. & Palmer, J. D. (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **348**, 337-339.

Douglas, S. E. (1994). Chloroplast origins and evolution. In *The Molecular Biology of Cyanobacteria*, pp. 91-118. Edited by D. A. Bryant. Dordrecht: Kluwer.

Douglas, S. E. & Penny, S. L. (1999). The plastid genome of the cryptophyte alga, *Guillardia theta*: Complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J. Mol. Evol.* **48**, 236-244.

Driessche, T. v. d., Bonotto, S. & Brachet, J. (1970). Inability of rifampicin to inhibit circadian rhythmicity in *Acetabularia* despite inhibition of RNA synthesis. *Biochem. Biophys. ACTA* **224**, 628-630.

Dron, M., Rahire, M. & Rochaix, J.-D. (1982). Sequence of the chloroplast 16S rRNA gene and its surrounding regions of *Chlamydomonas reinhardtii. Nucleic Acids Res.* **10**, 7609-7620.

Emes, M. J. & Tobin, A. K. (1993). Control of metabolism and development in higher plant plastids. *Int. Rev. Cytol.* **145**.

Erickson, J. M. (1996). Chloroplast transformation: Current results and future prospects. In *Oxygenic Photosynthesis: The Light Reactions* pp. 589-619. Edited by D. R. Ort & C. F. Yocum. Dordrecht: Kluwer.

Fong, S. E. & Surzycki, S. J. (1992). Chloroplast RNA polymerase genes of *Chlamydomonas reinhardtii* exhibit an unusual structure and arrangement. *Curr. Genet.* **21**, 485-497.

Franzén, L.-G. (1995). Analysis of chloroplast and mitochondrial targeting sequences from the green alga *Chlamydomonas reinhardtii. Membr. Cell Biol.* **8**, 315-322.

Gagne, G. & Guertin, M. (1992). The early genetic response to light in the green unicellular alga *Chlamydomonas eugametos* grown under light / dark cycles involves genes that represent direct responses to light and photosynthesis. *Plant Mol. Biol.* **18**, 429-445.

Goldschmidt-Clermont, M. (1991). Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker for site-directed transformation of *Chlamydomonas. Nucleic Acids Res.* **19**, 4083-4089.

Gorman, D. S. & Levine, R. P. (1965). Cytochrome F and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii. Proc. Natl. Acad. Sci. USA* **54,** 1665-1669.

Gray, M. W. (1989). The evolutionary origins of organelles. *Trends Genet.* **5**, 294-299.

Gray, M. W. & Lang, B. F. (1998). Transcription in chloroplasts and mitochondria: a tale of two polymerases. *Trends Microbiol.* **6**, 1-3.

Gruissem, W., Greenberg, B. M., Zurawski, G., Prescott, D. M. & Hallick, R. B. (1983). Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. *Cell* **35**, 815-828.

Gruissem, W., Narita, J. O., Greenberg, B. M., Prescott, D. M. & Hallick, R. B. (1983). Selective *in vitro* transcription of chloroplast genes. *J. Cell. Biochem.* **22**, 31-46.

Guertin, M. & Bellemare, G. (1979). Synthesis of chloroplast ribonucleic acid in *Chlamydomonas reinhardtii* toluene-treated cells. *Eur. J. Biochem.* **96**, 125-129.

Hajdukiewicz, P. T., Allison, L. A. & Maliga, P. (1997). The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* **16**, 4041-4048.

Hakimi, M.-A., Privat, I., Valay, J.-G. & Lerbs-Mache, S. (2000). Evolutionary conservation of C-terminal domains of primary sigma^{70}-type transcription factors between plants and bacteria. *J. Biol. Chem.* **275**, 9215-9221.

Hall, D. O. & Rao, K. K. (1995). *Photosynthesis.* Cambridge: Cambridge University Press.

Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* **21**, 3537-3544.

Hallick, R. B., Lipper, C., Richards, O. C. & Rutter, W. J. (1976). Isolation of a transcriptionally active chromosome from chloroplasts of *Euglena gracilis*. *Biochemistry* **15**, 3039-3045.

Harris, E. H. (1989). *The Chlamydomonas Sourcebook.* San Diego: Academic Press.

Hartmann, G., Honikel, K. O., Knusel, F. & Nuesch, J. (1967). The specific inhibition of the DNA-directed RNA synthesis by rifamycin. *Biochim. Biophys. Acta* **145**, 843-844.

Hedtke, B., Borner, T. & Weihe, A. (1997). Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. *Science* **277**, 809-811.

Heifetz, P. B. (2000). Genetic engineering of the chloroplast. *Biochimie* **82**, 655-666.

Heifetz, P. B., Forster, B., Osmond, C. B., Giles, L. J. & Boynton, J. E. (2000). Effects of acetate on facultative autotrophy in *Chlamydomonas reinhardtii* assessed by photosynthetic measurements and stable isotope analyses. *Plant Physiol.* **122**, 1439-1445.

Herrin, D. L. & Kuo, T.-C. (1998). RNA splicing in the chloroplast. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pp. 183-195. Edited by J.-D. Rochaix, M. Goldschmidt-Clermont & S. Merchant. Dordrecht: Kluwer.

Hess, W. R. & Borner, T. (1999). Organellar RNA polymerases of higher plants. *Int. Rev. Cytol.* **190**, 1-59.

Hess, W. R., Prombona, A., Fieder, B., Subramanian, A. R. & Borner, T. (1993). Chloroplast *rps*15 and the *rpo*B/C1/C2 gene cluster are strongly transcribed in ribosome-deficient plastids: evidence for a functioning non-chloroplast-encoded RNA polymerase. *EMBO J.* **12**, 563-571.

Hiejne, G. V., Steppuhn, J. & Herrmann, R. G. (1989). Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* **180**, 535-545.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura, M. (1989). The complete

sequence of the rice (Oryza sativa) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217**, 185-194.

Hodges, R. A., Perler, F. B., Noren, C. J. & Jack, W. E. (1992). Protein splicing removes intervening sequences in an archaea DNA polymerase. *Nucleic Acids Res.* **20**, 6153-6157.

Howe, C. J., Barbrook, A. C. & Lockhart, P. J. (2000). Organelle genes - do they jump or are they pushed? *Trends Genet.* **16**, 65-66.

Hu, J. & Bogorad, L. (1990). Maize chloroplast RNA polymerase: The 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. *Proc. Natl. Acad. Sci. USA* **87**, 1531-1535.

Hu, J., Troxler, R. F. & Bogorad, L. (1991). Maize chloroplast RNA polymerase: the 78-kilodalton polypeptide is encoded by the plastid *rpo*C1 gene. *Nucleic Acids Res.* **19**, 3431-3434.

Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M. & Liu, X. Q. (1994). The *Chlamydomonas* chloroplast *clp*P gene contains translated large insertion sequences and is essential for cell growth. *Mol. Gen. Genet.* **244**, 151-159.

Hudson, G. S., Holton, T. A., Whitfield, P. R. & Bottomley, W. (1988). Spinach chloroplast *rpo*BC genes encode three subunits of the chloroplast RNA polymerase. *J. Mol. Biol.* **200**, 639-654.

Ikeda, T. M. & Gray, M. W. (1999). Identification and characterisation of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. *Plant Mol. Biol.* **40**, 567-578.

Isono, K., Shimizu, M., Yoshimoto, K., Niwa, Y., Satoh, K., Yokota, A. & Kobayashi, H. (1997). Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of σ^{70} factors of bacterial RNA polymerases in *Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA* **94**, 14948-14953.

Jahn, D. (1992). Expression of the *Chlamydomonas reinhardtii* chloroplast $tRNA^{Glu}$ gene in a homologous *in vitro* transcription system is independent of upstream promoter elements. *Arch. Biochem. Biophys.* **298**, 505-513.

Jarvik, J. W. & Telmer C. A. (1998). Epitope tagging. *Annu. Rev. Genet.* **32**, 601-618.

Jennings, R. C. & Ohad, I. (1972). Biogenesis of chloroplast membranes. *Arch. Biochem. Biophys.* **153**, 79-87.

Jolly, S. O. & Bogorad, L. (1980). Preferential transcription of cloned maize chloroplast DNA sequences by maize chloroplast RNA polymerase. *Proc. Natl. Acad. Sci. USA* **77**, 822-826.

Kapoor, S., Suzuki, J. Y. & Sugiura, M. (1997). Identification and functional significance of a new class of non-consensus-type plastid promoters. *Plant J.* **11**, 327-337.

Kashlev, M., Martin, E., Polyakov, A., Severinov, K., Nikiforov, V. & Goldfarb, A. (1993). Histidine-tagged RNA polymerase: dissection of the transcription cycle using immobilized enzyme. *Gene* **130**, 9-14.

Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y. & Anraku, Y. (1997). Identification of three core regions essential for protein splicing of the yeast Vma1 protozyme. *J. Biol. Chem.* **272**, 15668-15674.

Kestermann, M., Neukirchen, S., Kloppstech, K. & Link, G. (1998). Sequence and expression characteristics of a nuclear-encoded chloroplast sigma factor from mustard (*Sinapsis alba*). *Nucleic Acids Res.* **26**, 2747-2753.

Kindle, K. L., Richards, K. L. & Stern, D. B. (1991). Engineering the chloroplast genome: Techniques and capabilities for chloroplast transformation in *Chlamydomonas reinhardtii. Proc. Natl. Acad. Sci. USA* **88**, 1721-1725.

Kirk, J. T. O. (1964). DNA-dependent RNA synthesis in chloroplast preparations. *Biochem. Biophys. Res. Commun.* **14**, 393-397.

Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., Palmer, J. D. & Roos, D. S. (1997). A plastid of probable green algal origin in apicomplexan parasites. *Science* **275**, 1485-1489.

Koonin, E. V. (1995). A protein splice-juction motif in hedgehog family proteins. *Trends Biochem. Sci.* **20**, 141-142.

Kowallik, K. V., Stoebe, B., Schaffran, I., Kroth-Pancic, P. & Freier, U. (1995). The chloroplast genome of a chlorophyll *a*+*c*-containing alga, *Odontella sinensis. Plant Mol. Biol. Rep.* **13**, 336-342.

Kozminski, K. G., Diener, D. R. & Rosenbaum, J. L. (1993). High level expression of nonacetylatable α-tubulin in *Chlamydomonas reinhardtii. Cell Motility and the Cytoskeleton* **25**, 158-170.

Krupinska, K. & Falk, J. (1994). Changes in RNA-polymerase activity during biogenesis, maturation and senescence of barley chloroplasts. Comparative analysis of transcripts synthesized either in run-on assays or by transcriptionally active chromosomes. *J. Plant Physiol.* **143**, 298-305.

Kruse, O., Nixon, P. J., Schmid, G. H. & Mullineaux, C. W. (1999). Isolation of state transition mutants of *Chlamydomonas reinhardtii* by fluorescence video imaging. *Photosynthesis Res.* **61**, 43-51.

Kung, S. D. & Lin, C. M. (1985). Chloroplast promoters from higher plants. *Nucleic Acids Res.* **13**, 7542-7543.

Kurnit, D. M. (1989). *Escherichia coli recA* deletion strains that are highly competent for transformation and for *in vivo* phage packaging. *Gene* **823,** 13-315.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lapidot, M., Raveh, D., Sivan, A., Arad, S. & Shapira, M. (1999). Cloning of the acetohydroxyacid synthase gene (AHAS) of *Porphyridium* sp. as a selection marker for chloroplast transformation. *2nd European Phycological Congress.* Montecatini Terme, Italy.

Lemieux, C., Otis, C. & Turmel, M. (2000). Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* **403**, 649-652.

Lerbs, S., Brautigam, E. & Mache, R. (1988). DNA-dependent RNA polymerase of spinach chloroplasts: Characterization of α-like and σ-like polypeptides. *Mol. Gen. Genet.* **211**, 459-464.

Lerbs, S., Brautigam, E. & Parthier, B. (1985). Polypeptides of DNA-dependent RNA polymerase of spinach chloroplasts: characterization by antibody-linked polymerase assay and determination of sites of synthesis. *EMBO J.* **4**, 1661-1666.

Lerbs-Mache, S. (1993). The 110-kDa polypeptide of spinach plastid DNA-dependent RNA polymerase: Single-subunit enzyme or catalyic core of multimeric enzyme complexes? *Proc. Natl. Acad. Sci. USA* **90**, 5509-5513.

Liere, K. & Maliga, P. (1999). *In vitro* characterisation of the tobacco *rpo*B promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters. *EMBO J.* **18**, 249-257.

Little, M. C. & Hallick, R. B. (1988). Chloroplast *rpo*A, *rpo*B and *rpo*C genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. *J. Biol. Chem.* **263**, 14302-14307.

Liu, B., Tan, S. & Troxler, R. F. (1999). Characterisation of a nuclear-encoded photogene for a chloroplast RNA polymerase sigma factor in the unicellular alga *Cyanidium caldarium* (Rhodophyta). *J. Phycol.* **35**, 778-785.

Liu, B. & Troxler, R. F. (1996). Molecular characterisation of a positively photoregulated nuclear gene for a chloroplast RNA polymerase σ factor in *Cyanidium caldarium*. *Proc. Natl. Acad. Sci. USA* **93**, 3313-3318.

Loiseaux, S., Mache, R. & Rozier, C. (1975). Rifampicin inhibition of the plastid rRNA synthesis of *Marchantia polymorpha*. *J. Cell Sci.* **17**, 327-335.

Lonetto, M., Gribskov, M. & Gross, C. A. (1992). The σ^{70} family: Sequence conservation and evolutionary relationships. *J. Bacteriol.* **174**, 3843-3849.

Lukens, J. H., Mathews, D. E. & Durbin, R. D. (1987). Effect of tagetitoxin on the levels of Ribulose 1,5-Bisphosphate Carboxylase, ribosomes, and RNA in plastids of wheat leaves. *Plant Physiol.* **84**, 808-813.

Maier, R. M., Neckermann, K., Igloi, G. L. & Kössel, H. (1995). Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* **251**, 614-628.

Maliga, P. (1998). Two plastid RNA polymerases of higher plants: an evolving story. *Trends Plant Sci.* **3**, 4-6.

Mandel, M. & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 159-162.

Martin, W. & Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: How much, what happens and why? *Plant Physiol.* **118**, 9-17.

Mathews, D. E. & Durbin, R. D. (1990). Tagetitoxin inhibits RNA synthesis directed by RNA polymerases from chloroplasts and *Escherichia coli. J. Biol. Chem.* **265**, 493-498.

Mathews, D. E. & Durbin, R. D. (1994). Mechanistic aspects of tagetitoxin inhibition of RNA polymerase from *Escherichia coli. Biochemistry* **33**, 11987-11992.

Mathys, S., Jr., T. C. E., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X.-Q. & Xu, M.-Q. (1999). Characterization of a self-splicing mini-intein and its

conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**, 1-13.

Matsumoto, T., Matsuo, M. & Matsuda, Y. (1991). Structural analysis and expression during dark-light transitions of a gene for cytochrome f in *Chlamydomonas reinhardtii. Plant Cell Physiol.* **32**, 863-872.

Morden, C. W., Wolfe, K. H., dePamphilis, C. W. & Palmer, J. D. (1991). Plastid translation and transcription genes in a non-photosynthetic plant: intact, missing and pseudo genes. *EMBO J.* **10**, 3281-3288.

Morikawa, K., Ito, S., Tsunoyama, Y., Nakahira, Y., Shiina, T. & Toyoshima, Y. (1999). Circadian-regulated expression of a nuclear-encoded plastid σ factor gene (*sigA*) in wheat seedlings. *FEBS Lett.* **451**, 275-275.

Narita, J. O., Rushlow, K. E. & Hallick, R. B. (1985). Characterization of a *Euglena gracilis* chloroplast RNA polymerase specific for ribosomal RNA genes. *J. Biol. Chem.* **260**, 11194-11199.

Neuhaus, H. E. & Emes, M. J. (2000). Non-photosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 111-140.

Newman, S. M., Gillham, N. W., Harris, E. H., Johnson, A. M. & Boynton J. E. (1991). Targeted disruption of chloroplast genes in *Chlamydomonas reinhardtii. Mol. Gen. Genet.* **230**, 65-74.

Newman, S. M., Boynton, J. E., Gillham, N. W., Randolph-Anderson, B. L., Johnson, A. M. & Harris, E. H. (1990). Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: Molecular and genetic characterisation of integration events. *Genetics* **126**, 875-888.

Nickelsen, J., Fleischmann, M., Boudreau, E., Rahire, M. & Rochaix, J.-D. (1999). Identification of *cis*-acting RNA leader elements required for chloroplast *psbD* gene expression in *Chlamydomonas. Plant Cell* **11**, 957-970.

Nickelsen, J. & Kuck, U. (2000). The unicellular green alga *Chlamydomonas reinhardtii* as an experimental system to study chloroplast RNA metabolism. *Naturwissenschaften* **87**, 97-107.

Nugent, J. H. A. (1996). Oxygenic photosynthesis. Electron transfer in photosystem I and photosystem II. *Eur. J. Biochem.* **237**, 519-531.

O'Connor, H. E., Ruffle, S. V., Cain, A. J., Deak, Z., Vass, I., Nugent, J. H. A. & Purton, S. (1998). The 9-kDa phosphoprotein of photosystem II. Generation and characterisation of *Chlamydomonas* mutants lacking PSII-H and a site-directed mutant lacking the phosphorylation site. *Biochem. Biophys. Acta* **1364,** 63-72.

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 organisation deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **322**, 572-574.

Oikawa, K., Tanaka, K. & Takahashi, H. (1998). Two types of differentially photo-regulated nuclear genes that encode s factors for chloroplast RNA polymerase in the red alga *Cyanidium caldarium* strain RK-1. *Gene* **210**, 277-285.

Olive, J. & Wollman, F.-A. (1998). Supramolecular organization of chloroplast and thylakoid membranes. In *The Molecular Biology of*

Chloroplasts and Mitochondria in Chlamydomonas. Edited by J.-D. Rochaix, M. Goldschmidt-Clermont & S. Merchant. Dordrecht: Kluwer.

Palmer, J. D. (1997). Organelle genomes: Going, going, gone! *Science* **275**, 790-791.

Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. & Belfort, M. (1994). Protein splicing elements: inteins and exteins - a definition of terms and recommended nomenclature. *Nucleic Acids Res.* **22**, 1125-1127.

Pfannschmidt, T. & Link, G. (1994). Separation of two classes of plastid DNA-dependent RNA polymerases that are differentially expressed in mustard (*Sinapis alba* L.) seedlings. *Plant Mol. Biol.* **25**, 69-81.

Pfannschmidt, T. & Link, G. (1997). The A and B forms of plastid DNA-dependent RNA polymerase from mustard (*Sinapis alba* L.) transcribe the same genes in a different developmental context. *Mol. Gen. Genet.* **257**.

Pfannschmidt, T., Ogrzewalla, K., Baginsky, S., Sickmann, A., Meyer, H. E. & Link, G. (2000). The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba* L.). *Eur. J. Biochem.* **267**, 253-261.

Polya, G. M. & Jagendorf, A. T. (1971). Wheat leaf RNA polymerases. 1. partial purification and characterization of nuclear, chloroplast and soluble DNA-dependent enzymes. *Arch. Biochem. Biophys.* **146**, 635-648.

Porra, R. J., Thompson, W. A. & Kriedemann, P. E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: verification of the

concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochem. Biophys. Acta* **975**, 384-394.

Purton, S. (1995). The chloroplast genome of *Chlamydomonas. Science Progress* **78**, 205-216.

Purton, S. & Rochaix, J.-D. (1995). Characterisation of the ARG7 gene of *Chlamydomonas reinhardtii* and its application to nuclear transformation. *Eur. J. Phycol.* **30**, 141-148.

Purton, S. & Gray, J. C. (1989). The plastid *rpo*A gene encoding a protein homologous to the bacterial RNA polymerase alpha subunit is expressed in pea chloroplasts. *Mol. Gen. Genet.* **217**, 77-84.

Race, H. L., Herrmann, R. G. & Martin, W. (1999). Why have organelles retained genomes? *Trends Genet.* **15**, 364-370.

Reardon, E. M. & Price, C. A. (1995). Plastid genomes of three non-green algae are sequenced. *Plant Mol. Biol. Rep.* **13**, 320-326.

Reiss, T. & Link, G. (1985). Characterization of transcriptionally active DNA-protein complexes from chloroplasts and etioplasts of mustard (*Sinapis alba* L.). *Eur. J. Biochem.* **148**, 207-212.

Reith, M. & Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Rep.* **13**, 333-335.

Rochaix, J.-D. (1987). Molecular genetics of chloroplasts and mitochondria in the unicellular green alga *Chlamydomonas. FEMS Microbiology Reviews* **46**, 13-34.

Rochaix, J.-D., Mayfield, S., Goldschmidt-Clermont, M. & Erickson, J. (1988). Molecular biology of *Chlamydomonas*. In Plant Molecular Biology, pp. 253-257. Edited by J. N. Shaw.

Rochaix, J.-D. (1992). Post-transcriptional steps in the expression of chloroplast genes. *Annu. Rev. Cell Biol.* **8**, 1-28.

Rochaix, J.-D. (1995). *Chlamydomonas reinhardtii* as the photosynthetic yeast. *Annu. Rev. Genet.* **29**, 209-230.

Rochaix, J.-D. (1997). Chloroplast reverse genetics: new insights into the function of plastid genes. *Trends Plant Sci.* **2**, 419-425.

Ruf, M. & Kössel, H. (1988). Structure and expression of the gene coding for the α-subunit of DNA-dependent RNA polymerase from the chloroplast genome of *Zea mays*. *Nucleic Acids Res.* **16**, 5741-5754.

Rushlow, K. E., Jr., E. M. O., Lipper, C. & Hallick, R. B. (1980). Selective *in vitro* transcription of *Euglena* chloroplast ribosomal RNA genes by a transcriptionally active chromosome. *J. Biol. Chem.* **255**, 3786-3792.

Sakai, A., Saito, C., Inada, N. & Kuroiwa, T. (1998). Transcriptional activities of the chloroplast-nuclei and proplastid-nuclei isolated from tobacco exhibit different sensitivities to tagetitoxin: Implication of the presence of distinct RNA polymerases. *Plant Cell Physiol.* **39**, 928-34.

Sakamoto, W., Kindle, K. L. & Stern, D. B. (1993). *In vivo* analysis of *Chlamydomonas* chloroplast gene expression using stable transformation of β-glucuronidase translational fusions. *Proc. Natl. Acad. Sci. USA* **90**, 497-501

Sambrook, J., Fritsch, E. F. & Maniatis, T (1989) *Molecular Biology: A Laboratory Manual.* New York: Cold Spring Harbour Laboratory Press

Santis-Maciossek, G. D., Kofer, W., Bock, A., Schoch, S., Maier, R. M., Wanner, G., Rudiger, W., Koop, H.-U. & Herrmann, R. G. (1999). Targeted disruption of the plastid RNA polymerase genes *rpo*A, B and C1: molecular biology, biochemistry and ultrastructure. *Plant J.* **18**, 477-489.

Sato, S., Tews, I. & Wilson, R. J. M. (2000). Impact of a plastid-bearing endocytobiont on apicomplexan genomes. *Int. J. Parasitology* **30**, 427-439.

Schleif, R. (1969). Isolation and characterisation of a streptolydigin resistant RNA polymerase. *Nature* **233**, 1068-1069.

Schneider, M., Darlix, J-L., Erikson, J. & Rochaix, J.-D. (1985). Sequence organization of repetitive elements in the flanking regions of the chloroplast ribosomal unit of *Chlamydomonas reinhardtii. Nucleic Acids Res.* **13**, 8529-8541.

Sears, B. B. (1998). Chloroplast replication, recombination and repair. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pp. 115-138. Edited by J.-D. Rochaix, M. Goldschmidt-Clermont & S. Merchant. Dordrecht: Kluwer.

Serino, G. & Maliga, P. (1998). RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiol.* **117**, 1165-1170.

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.

Sijben-Muller, G., Hallick, R. B., Alt, J., Westhoff, P. & Herrmann, R. G. (1986). Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase α-subunit. *Nucleic Acids Res.* **14**, 1029-1044.

Silflow, C. D. (1998). Organization of the nuclear genome. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pp. 115-138. Edited by J.-D. Rochaix, M. Goldschmidt-Clermont & S. Merchant. Dordrecht: Kluwer.

Silhavy, D. & Maliga, P. (1998). Mapping of promoters for the nucleus-encoded plastid RNA polymerase (NEP) in the *iojap* maize mutant. *Curr. Genet* **33**, 340-344.

Snustad, D. P., Simmons, M. J. & Jenkins, J. B. (1997). *Principles of genetics.* New York: John Wiley & Sons, Inc.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Spencer, D. & Whitfield, P. R. (1967). Ribonucleic acid synthesizing activity of spinach chloroplasts and nuclei. *Arch. Biochem. Biophys.* **121**, 336-345.

Stanley, K. K. & Luzio, J. P. (1984). Construction of a new family of high efficiency bacterial expression vectors: Identification of cDNA clones coding for human liver proteins. *EMBO J.* **3**, 1429-1434.

Stern, D. B. & Drager, R. G. (1998). Chloroplast transcription. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pp. 165-181. Edited by J.-D. Rochaix, M. Goldschmidt-Clermont & S. Merchant. Dordrecht: Kluwer.

Stevens, D. R. & Purton, S. (1997). Genetic engineering of eukaryotic algae: Progress and prospects. *J. Phycol.* **33**, 713-722.

Stirewalt, V. L., Michalowski, C. B., Loffelhardt, W., Bohnert, H. J. & Bryant, D. A. (1995). Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol. Biol. Rep.* **13**, 327-332.

Stoebe, B., Martin, W. & Kowallik, K. V. (1998). Distribution and nomenclature of protein-coding genes in twelve sequenced chloroplast genomes. *Plant Mol. Biol. Rep.* **16**, 243-255.

Suck, R., Zeltz, P., Falk, J., Acker, A., Kössel, H. & Krupinska, K. (1996). Transcriptionally active chromosomes (TACs) of barley chloroplasts contain the α-subunit of plastome-encoded RNA polymerase. *Curr. Genet.* **30**, 515-521.

Sugiura, M. (1995). The chloroplast genome. *Essays Biochem.* **30**, 49-57.

Sugiura, M., Inoue, Y. & Minagawa, J. (1998). Rapid and discrete isolation of oxygen-evolving His-tagged photosystem II core complex from *Chlamydomonas reinhardtii* by Ni^{2+} affinity column chromatography. *FEBS Lett.* **426**, 140-144.

Sulkowski, E. (1985). Purification of proteins by IMAC. *Trends Biotech.* **3**, 1-7.

Surzycki, S. J. (1969). Genetic functions of the chloroplast of *Chlamydomonas reinhardtii*: Effect of rifampin on chloroplast DNA-dependent RNA polymerase. *Biochemistry* **63**, 1327-1334.

Surzycki, S. J. & Shellenbarger, D. L. (1976). Purification and characterization of a putative sigma factor from *Chlamydomonas reinhardti*. *Proc. Natl. Acad. Sci. USA* **73**, 3961-3965.

Svab, Z. & Maliga, P. (1993). High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* **90**, 913-917.

Tan, S. & Troxler, R. F. (1999). Characterization of two chloroplast RNA polymerase sigma factors from *Zea mays*: Photoregulation and differential expression. *Proc. Natl. Acad. Sci. USA* **96**, 5316-5321.

Tanaka, K., Oikawa, K., Ohta, N., Kuroiwa, H., Kuroiwa, T. & Takahashi, H. (1996). Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. *Science* **272**, 1932-1935.

Tanaka, K., Tozawa, Y., Mochizuki, N., Shinozaki, K., Nagatani, A., Wakasa, K. & Takahashi, H. (1997). Characterisation of three cDNA species encoding plastid RNA polymerase sigma factors in *Arabidopsis thaliana*: evidence for the sigma factor heterogeneity in higher plant plastids. *FEBS Lett.* **413**, 309-313.

Tewari, K. K. & Goel, A. (1983). Solubilization and partial purification of RNA polymerase from pea chloroplasts. *Biochemistry* **22**, 2142-2148.

Tewari, K. K. & Wildman, S. G. (1969). Function of chloroplast DNA. *Biochem. Biophys. Acta* **186**, 358-372.

Tiller, K., Eisermann, A. & Link, G. (1991). The chloroplast transcription apparatus from mustard (*Sinapis alba* L.). Evidence for three different transcription factors which resemble bacterial σ factors. *Eur. J. Biochem.* **198**, 93-99.

Tiller, K. & Link, G. (1993). Phosphorylation and dephosphorylation affect functional characteristics of chloroplast and etioplast transcription systems from mustard (*Sinapsis alba* L.). *EMBO J.* **12**, 1745-1753.

Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.

Tozawa, Y., Tanaka, K., Takahashi, H. & Wakasa, K. (1998). Nuclear encoding of a plastid σ factor in rice and its tissue- and light-dependent expression. *Nucleic Acids Res.* **26**, 415-419.

Troxler, R. F., Zhang, F., Hu, J. & Bogorad, L. (1994). Evidence that σ factors are components of chloroplast RNA polymerase. *Plant Physiol.* **104**, 753-759.

Turmel, M. & Otis, C. (1994). The chloroplast gene cluster containing *psb*F, *psb*L, *pet*G and *rps*3 is conserved in *Chlamydomonas*. *Curr. Genet.* **27**, 54-61.

Turmel, M., Otis, C. & Lemieux, C. (1999). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the

architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. USA* **96**, 10248-10253.

Tyers, M., Tokiwa, G., Nash, R. & Flutcher, B. (1992). The Cln3-Cdc28 kinase complex of *S. cerevisiae* is regulated by proteolysis and phosphorylation. *EMBO J.* **11**, 1773-1784.

von Hiejne, G., Steppuhn, J. & Herrmann, R. G. (1989). Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* **180**, 535-545.

Wakasugi, T., Nagai, T., Kspoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K. & Sugiura, M. (1997). Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. USA* **94**, 5967-5972.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T. & Sugiura, M. (1994). Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii. Proc. Natl. Acad. Sci. USA* **91**, 9794-9798.

Wang, S. & Liu, X. Q. (1997). Identification of an unusual intein in chloroplast ClpP protease of *Chlamydomonas eugametos. J. Biol. Chem.* **272**, 11869-11873.

Weihe, A. & Borner, T. (1999). Transcription and the architecture of promoters in chloroplasts. *Trends Plant Sci.* **4**, 169-170.

Wilson, R. J. M., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W. & Williamson, D. H. (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparium*. *J. Mol. Biol.* **261**, 155-172.

Wolfe, K. H., Morden, C. W. & Palmer, J. D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. USA* **89**, 10648-10652.

Xu, M. Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J. & Perler, F. B. (1993). *In vitro* protein splicing of purified precursor and the identification of a branched intermediate. *Cell* **75**, 1371-1377.