Tools for nuclear transformation in

Chlamydomonas reinhardtii

A thesis submitted for the degree of
Doctor of Philosophy
by

Kathryn Louise Madagan B. Sc. (Hons)

Department of Biology
University College London

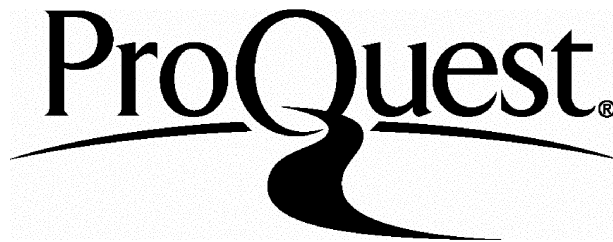ProQuest Number: U642186

Pro**Q**uest.

ProQuest U642186

Thesis Abstract

Due to its simple life cycle and classical Mendelian genetics the green alga *Chlamydomonas reinhardtii* has long been used as a model system for the study of flagellar function, photosynthesis and response to light. Transformation of the nuclear genome has now become routine and with this the development of a variety of promoters, transformation markers and methods for genomic complementation and insertional mutagenesis has taken place. However, in order to exploit fully the potential of *Chlamydomonas*, further procedures for molecular genetic investigation are required. For example, it is desirable to develop a detailed molecular map of the *Chlamydomonas* nuclear genome. Also there is a need to find a suitable reporter molecule that can act as a marker for gene expression and protein localisation. Other desirable advances would include a way of targeting specific genes by selection for rare homologous recombination events and a means for down-regulation of genes by the application of antisense technology.

This thesis presents the results of investigation into possible new tools for nuclear transformation. Firstly, the results of a pilot study of analysis of Expressed Sequence Tags from *Chlamydomonas* are presented. Sixteen cDNAs chosen at random from a *Chlamydomonas* cDNA library were sequenced. The sequences were analysed for sequence similarities to other organisms and for functional motifs. Secondly, the use of Green Fluorescent Protein as a reporter molecule for *Chlamydomonas* was investigated. The GFP coding sequence under control of a *Chlamydomonas* promoter was introduced into the nuclear genome and the transformants examined for expression of Green Fluorescent Protein. Thirdly attempts were made to develop a negative selectable marker for homologous transformation based on the creation of mutants defective in uracil or acetate utilisation and their restoration to wild-type phenotype by transformation with the wild-type gene. Finally the use of antisense expression was investigated. An antisense section of the *Chlamydomonas* nuclear gene encoding oxygen evolving complex 1 was inserted into the genome of *Chlamydomonas* under the control of a strong *Chlamydomonas* promoter and the resulting transformants analysed for phenotypic effects and the presence of an antisense transcript.

For Mum and Dad,
with love.

Acknowledgments

Abbreviations

AMPS	ammonium persulphate
BAC	bacterial artificial chromosome
BEAUTY	BLAST enhanced alignment utility
BLAST	basic local alignment search tool
BSA	bovine serum albumin
bp (kbp)	base pairs (kilobase pairs)
CaMv	cauliflower mosaic virus
cDNA	copy deoxyribonucleic acid
Da (kDa)	dalton (kilodalton)
dATP (A)	2' deoxyadenosine 5'-triphosphate
dCTP (C)	2' deoxycytidine 5'-triphosphate
dGTP (G)	2' deoxyguanosine 5'-triphosphate
dNTP	2' deoxynucleoside 5'-triphosphate
dTTP (T)	2' deoxythymidine 5'-triphosphate
ddATP	2',3' dideoxyadenosine 5'-triphosphate
ddCTP	2',3' dideoxycytidine 5'-triphosphate
ddH2O	double distilled water
ddGTP	2',3' dideoxyguanosine 5'-triphosphate
ddNTP	2',3' dideoxynucleoside 5'-triphosphate
ddTTP	2',3' dideoxythymidine 5'-triphosphate
DEAE	diethylaminoethyl
DDJB	DNA database of Japan
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA.Na$_2$	diaminoethanetetra-acetic acid. disodium salt
EGFP	enhanced green fluorescent protein
ELISA	enzyme linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
EST	expressed sequence tag
EtBr	ethidium bromide
FAM	fluoro-acetamide
FOA	fluoro-orotic acid
G418	geneticin
GANC	gancyclovir

GFP	green fluorescent protein
GUS	β-glucoronidase
HSP	high-scoring segment pairs
HSV-*tk*	Herpes Simplex virus
IPTG	isopropylthio-b-galactoside
LB (LBA)	Luria broth medium (+agar)
MOPS	3-[N-morpholino]propanesulphonic acid
Mb	megabases
mRNA	messenger ribonucleic acid
mt	mating type
OEE	oxygen evolving enhancer
OMP	orotidine-5'-monophosphate
NCBI	National Center for Biotechnology Information
Neo	neomycin
PAC	phage artificial chromosome
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PG	polygalacuronase
PIR	Protein Identification Resources
PMSF	phenylmethylsulphonyl fluoride
RACE	rapid amplification of cDNA ends
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SSPE	saline sodium phosphate-EDTA
SSC	saline sodium citrate
STS	sequence tagged site
TAP	tris acetate phosphate medium
TBE	tris borate-EDTA
TE	Tris-EDTA
TEMED	*N,N,N',N* - tetramethylethylenediamine
TEN	Tris-EDTA saline
Tricine	tris(hydroxymethyl)methylglycine
Tris	tris(hydroxymethyl)aminoethane
Tris min	Tris minimal medium

TTBS	Tween tris buffered saline
tRNA	transfer ribonucleic acid
UMP	uridine-5'-monophosphate
UTR	untranslated region
UV	ultra-violet radiation
v/v	volume for volume
wt	wild-type
w/v	weight for volume
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactoside
XSO_4	5-bromo-4-chloro-3-indolyl-sulphate
YAC	yeast artificial chromosome

Table of Contents

List of Figures

11

List of Tables

Chapter 1
INTRODUCTION

Chapter 1: Introduction

1.1 *Chlamydomonas reinhardtii* - a model system

1.1.1 History and background

Chlamydomonas is a eukaryotic, unicellular green alga that can photosynthesise, is motile and displays a phototactic response. *Chlamydomonas* has been used for studies of subjects as diverse as photosynthesis, flagellar assembly and function, mating reactions and gametogenesis, genetics and biosynthesis of organelles, cell wall synthesis, behavioural response to light, circadian rhythms and carbon, nitrogen and sulphur metabolism. The alga has well defined genetics and provides a powerful model system for the study of this wide range of cellular processes.

The genus *Chlamydomonas* contains several hundred described species and has a wide distribution, from marine and freshwater environments to soil and the atmosphere. Three species of *Chlamydomonas, C. eugametos, C. moewusii* and *C. reinhardtii* have historically been used in genetic and cytological studies. *Chlamydomonas reinhardtii* was first isolated in 1876 (Reinhardt 1876). Interest in the organism for research dates back to early cytological observations of haploid chromosome numbers (Dangeard 1888) and genetic studies in which tetrad analysis was carried out for the first time for any organism (Pascher 1918). This thesis deals exclusively with *C. reinhardtii.*

1.1.2 The physical features of *Chlamydomonas*

Wild-type cells of *Chlamydomonas* are oval-shaped, typically 10μm in length and 3μm in width, possessing two flagella at their anterior end (figure 1.1). The nucleus is usually centrally located with a prominent nucleolus and the cell contains a single cup-shaped chloroplast which may partially surround the nucleus. The chloroplast may occupy 40-60% of total cell volume and houses a photosynthetic apparatus closely resembling that of higher plants. One or more pyrenoids are found within the chloroplast, and starch bodies are often seen surrounding these pyrenoids. Also present are several smaller mitochondria occupying 1-3% of total cell volume, and a coloured eyespot which uses rhodopsin as the photoreceptor. The species has a cell wall composed from proline-rich glycoproteins, and a contractile vacuole. These features are all apparent by light microscopy under good conditions and have been the basis for species recognition.

1.1.3 The genetic systems of *Chlamydomonas*

Like other photosynthetic eukaryotes, *Chlamydomonas* possesses three distinct genomes, each located in a separate organelle. The bulk of total DNA is in the cell nucleus and the majority of proteins found in mitochondria and plastids are encoded in the nuclear genome.
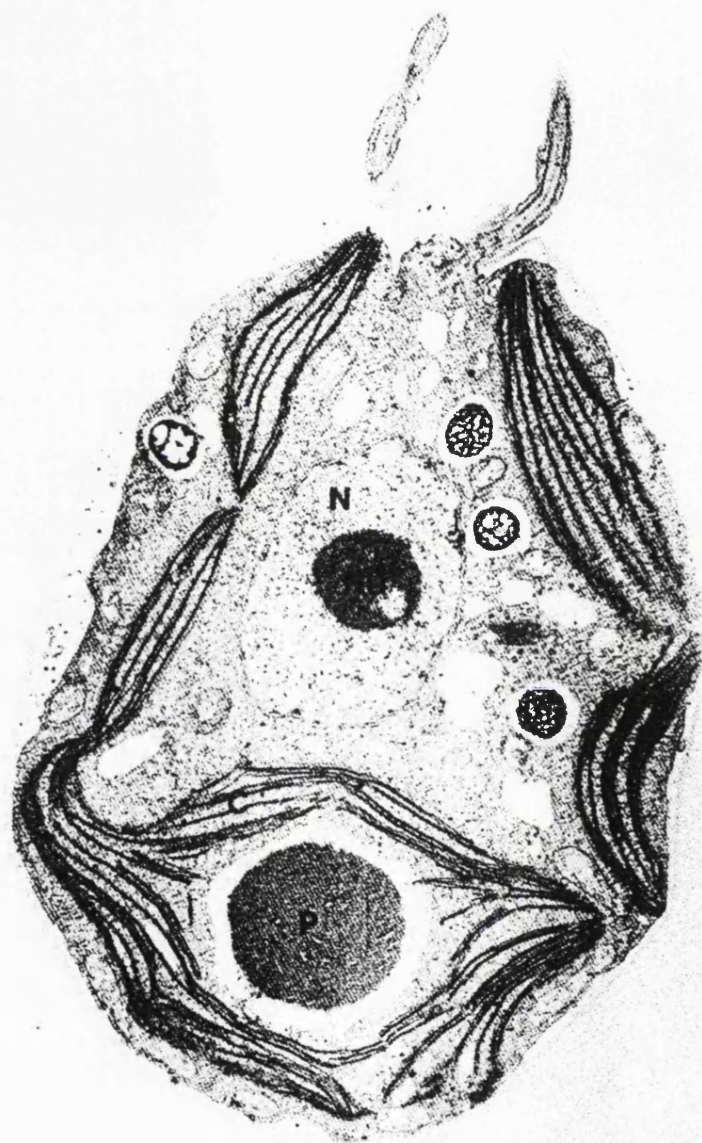
Figure 1.1 *Electron micrograph of wild-type* Chlamydomonas reinhardtii *cell showing prominent nucleus* (**N**) *and nucleolus* (**N U**), *chloroplast* (**C**), *pyrenoid* (**P**), *mitochondria* (**M**) *and flagellum* (**F**). *Courtesy of J. E. Boynton.*

16

However, both the mitochondrial and plastid genomes contain significant amounts of DNA. This non-nuclear DNA is replicated and expressed within those organelles, and the gene products are essential for their function. The chloroplast DNA of *Chlamydomonas* consists of 196 kbp circular molecules, present in about 80 copies per cell. The mitochondrial genome consists of 15.8 kbp linear molecules and is also present in about 80 copies per cell.

Nuclear DNA Organisation

Chlamydomonas is haploid and its nuclear genome is relatively small, consisting of approximately 10^8 base pairs, encoding some 10 000 genes (Harris 1989). The genetic map of the nucleus is composed of 148 loci spread over 17 linkage groups (figure 1.2). The markers include auxotrophic mutations, drug resistance markers, mutations affecting photosynthesis, flagellar function, and mating. The construction of RFLP maps for several nuclear linkage groups is in progress (Ranum, Thompson et al. 1988; Silflow, Kathir et al. 1995). Currently, the segregation patterns for approximately 200 molecular markers have been analysed using mostly RFLP markers and some short tagged sequence (STS) markers. The markers have been mapped to all 17 of the known linkage groups, with an average spacing of 4-5 cM or 0.4-0.5 megabases (Rochaix 1995).

A strong codon bias has been observed for *Chlamydomonas* nuclear genes that are highly expressed e.g. those encoding proteins of the photosynthetic and flagellar apparatus (Harris 1989) (see Table 1 a). Cytosine or guanine bases at the third position of the codons and if possible in the first position are preferred, reflecting the high G+C content of the nuclear DNA. A considerably more balanced codon usage is observed for genes that have a lower rate of expression e.g. cytochrome c_6 (Hill, Li et al. 1991) and aryl-sulphatase (de Hostos, Schilling et al. 1989). Adenine-rich sequence elements resembling the TATA box are usually present upstream of the nuclear genes with an adjacent G+C rich stretch. A putative polyadenylation recognition motif TGTAA is found in all cases examined at the 3' end, 10-15 nucleotides upstream of the polyadenylation site. Another characteristic feature of *Chlamydomonas* nuclear genes is the presence of multiple introns, whereas their higher plant homologues often have fewer introns or no introns at all (Muller, Igloi et al. 1992).

The chloroplast genome

It is well established that chloroplasts contain their own genetic system comprising DNA, RNA, DNA- and RNA-polymerases, ribosomes and translation factors. This system co-operates closely with its counterpart in the nuclear-cytosol compartment in the biosynthesis of chloroplast components. Most prominent among these are the photosynthetic apparatus and the chloroplast protein synthesising system (Mayfield, Bennoun et al. 1987a).

The chloroplast genome of *Chlamydomonas* has not yet been completely sequenced, but a complete restriction map has been published (Harris 1989). The DNA sequence data exists

LINKAGE GROUP

I pf-22 sup-2 y-6 arg-1 arg-7 pab-2 ac-20 msr-1 thi-3
ac-14 ac-115 ac-24 pf-4
y-5, y-10, pc-1, ery-3
su-4, ac-211, cr-7 shf-2

II fsr-1 ac-40 pf-21 ac-12 lf-1 act-1 ac-80 nic-2
thi-9 pf-12 pf-18

III y-6 ac-28 nit-2 y-7 ac-17 ac-141 ac-210 thi-2
pf-15 pf-5 ac-26 ac-207
pab-1
ac-208

IV nic-11 pyr-1 pf-20 ac-55
thi-4

V ac-31 ac-18 pf-1
thi-8 pf-26

VI fla-6 mt° mbo-1 shf-1 vfl-3 pf-14 ac-30 act-2
ac-212 ac-32 sup-3

*MATING TYPE
CLUSTER: mt, ac-29,
ac-213, cyt-1, fa-1, imp-1,
mat-2, mat-3, nic-7,
sad-1, thi-10, imp-11

VII ac-6 pf-17 ao-1 ac-22 ac-5
can-1

10% RECOMBINATION

VIII nr-1 pf-3 thi-1 vfl-1 ac-157

IX ac-51 sup-1 pf-16 ac-15 nit-1 pf-13
smr-1 sr-1 mbo-2

X fla-8 ac-16 ac-145 nic-13 fla-4 pr-1 pf-25
pf-19 ey-1 fla-3
pf-24
pf-6

LINKAGE GROUP

XI pf-23 w-7 fla-6 pf-2 ac-7 ery-1 ac-21
es-1 ac-35

XII lf-2 arg-4 ac-69 ac-9 pf-27 ac214 gln-1

XIII pf-9 cr-1 ac-19 nic-15 mbo-3

XIV ac-206 ac-177 agg-1

XV nic-1

10% RECOMBINATION

XVI ac-46

XVII y-1

XVIII ani-1 spr-1 ac-13 ac-215

UNI
(CIRCULAR)
uni-1 fla-9 fla-10 pf-10 pf-29 sun-1 fla-f1 uni-1
fla-13 fla-f2
–enh-1 –
pf-7 pf-8

Figure 1.2 *Linkage map of* Chlamydomonas reinhardtii
nuclear genome. From Harris et al. 1987.

*until recently it was thought that there were 19 linkage groups for *Chlamydomonas*, but it has been found that
groups 12/13 and 16/17 are linked to one another. Source: ChlamyDB.

	Nu	Ct		Nu	Ct		Nu	Ct		Nu	Ct
Phe			**Ser**			**Tyr**			**Cys**		
UUU	1.6	21.5	UCU	7.2	36	UAU	0	18	UGU	0	96.5
UUC	98.4	78.5	UCC	43.9	2	UAC	100	32	UGC	100	3.5
			UCA	0.6	45						
Leu			UCG	33.9	2	**His**			**Trp**		
UUA	0	58.5	AGU	0	9	CAU	4.4	6	UGG	100	100
UUG	0	1.5	AGC	14.4	5	CAC	95.6	94			
CUU	1.0	27.0							**Arg**		
CUC	4.0	0	**Pro**			**Gln**			CGU	10.2	96.0
CUA	0.5	13.0	CCU	3.2	38	CAA	0	87.5	CGC	89.0	3.0
CUG	94.5	0	CCC	87.9	0	CAG	100	12.5	CGA	0	0
			CCA	0	57.5				CGG	0.8	0
Ile			CCG	8.9	4.5	**Asn**			AGA	0	1.0
AUU	16.5	60				AAU	1.9	3.5	AGG	0	0
AUC	83.5	36	**Thr**			AAC	98.1	96.5			
AUA	0	4	ACU	9.5	59						
			ACC	89.1	0	**Lys**			**Gly**		
Met			ACA	0	39	AAA	0	95.0	GGU	22.0	95.5
AUG	100	100	ACG	1.4	2	AAG	100	5.0	GGC	76.6	4
									GGA	0	0.5
Val			**Ala**			**Asp**			GGG	0.4	0
GUU	3.8	45.5	GCU	18.0	64.5	GAU	16.1	31.0			
GUC	44.0	0	GCC	76.2	0.5	GAC	83.9	69.0	**Stop**		
GUA	0	54.5	GCA	0	19.5				UAA	100	100
GUG	52.2	0	GCG	5.8	15.5	**Glu**			UAG	0	0
						GAA	87.0	87.0	UGA	0	0
						GAG	13.0	13.0			

Table 1 a   *Codon usage table for Chlamydomonas nuclear and chloroplast genomes*

for about 75% of the genome, with sequence analysis for the remaining regions currently underway or planned. Approximately 80 genes have so far been identified (Purton 1995), including 26 genes encoding photosynthetic proteins, 18 tRNA genes, 14 ribosomal protein genes, five ribosomal RNA genes, three genes encoding subunits of RNA polymerase and several open reading frames of unknown function that are also present in higher plants (figure 1.3).

The chloroplast DNA of *Chlamydomonas* consists of 196 kbp circular molecules present in about 80 copies per cell. The DNA is organised into eight to ten DNA-protein complexes, called nucleoids. The DNA is A+T rich and therefore is at a lower density than that of nuclear DNA. The strikingly different codon usage in chloroplast and nuclear genes suggests that these genomes have different origins, in agreement with the endosymbiotic theory. This theory proposes that chloroplasts have evolved from photosynthetic prokaryotes that invaded and established an internal symbiotic relationship with the eukaryotic ancestor of plants and green algae (Gray 1992). Chloroplast genomes share many features common to prokaryotic genomes, including the presence of polycistrons, rRNA binding sequences in the 5' untranslated region of the mRNA and stem-loop structures near the 3' end of the mRNAs. Chloroplast gene expression is usually regulated by nuclear encoded regulatory factors (Rochaix 1992).

Mitochondrial DNA organisation

Like the chloroplast, the mitochondrion also has its own genetic system. Analysis of the *Chlamydomonas* mitochondrial genome has revealed it to consist of a 15.8 kbp linear molecule with inverted repeats at its ends and, similar to the chloroplast genome, present in multiple copies per cell (Harris 1989). Sequencing of this mitochondrial genome has shown that it contains eight protein genes, three tRNA genes and two ribosomal RNA genes (Boer, Bonen et al. 1985; Boer and Gray 1988; Michaelis, Varenholz et al. 1990) (figure 1.4).

1.1.4 The use of *Chlamydomonas* as a model organism

Chlamydomonas has a number of features that make it the ideal model system. The alga can be maintained either in liquid culture or on agar plates as individual colonies. It has a short generation time and the growth medium is simple and cheap to produce. Cells are grown phototrophically on a simple mineral medium with CO_2 as the carbon source in the light, or with acetate as a source of reduced carbon in the dark (heterotrophy) or in light (mixotrophy). The cell division cycle of the alga can also be synchronised by subjecting the cells to alternate light and dark cycles and a single life cycle may be accomplished in 7-10 hours (Harris 1989).

Ease of culture and the relatively small genome of the organism allow the rapid generation of interesting mutants. The organism is haploid and has a simple sexual cycle so
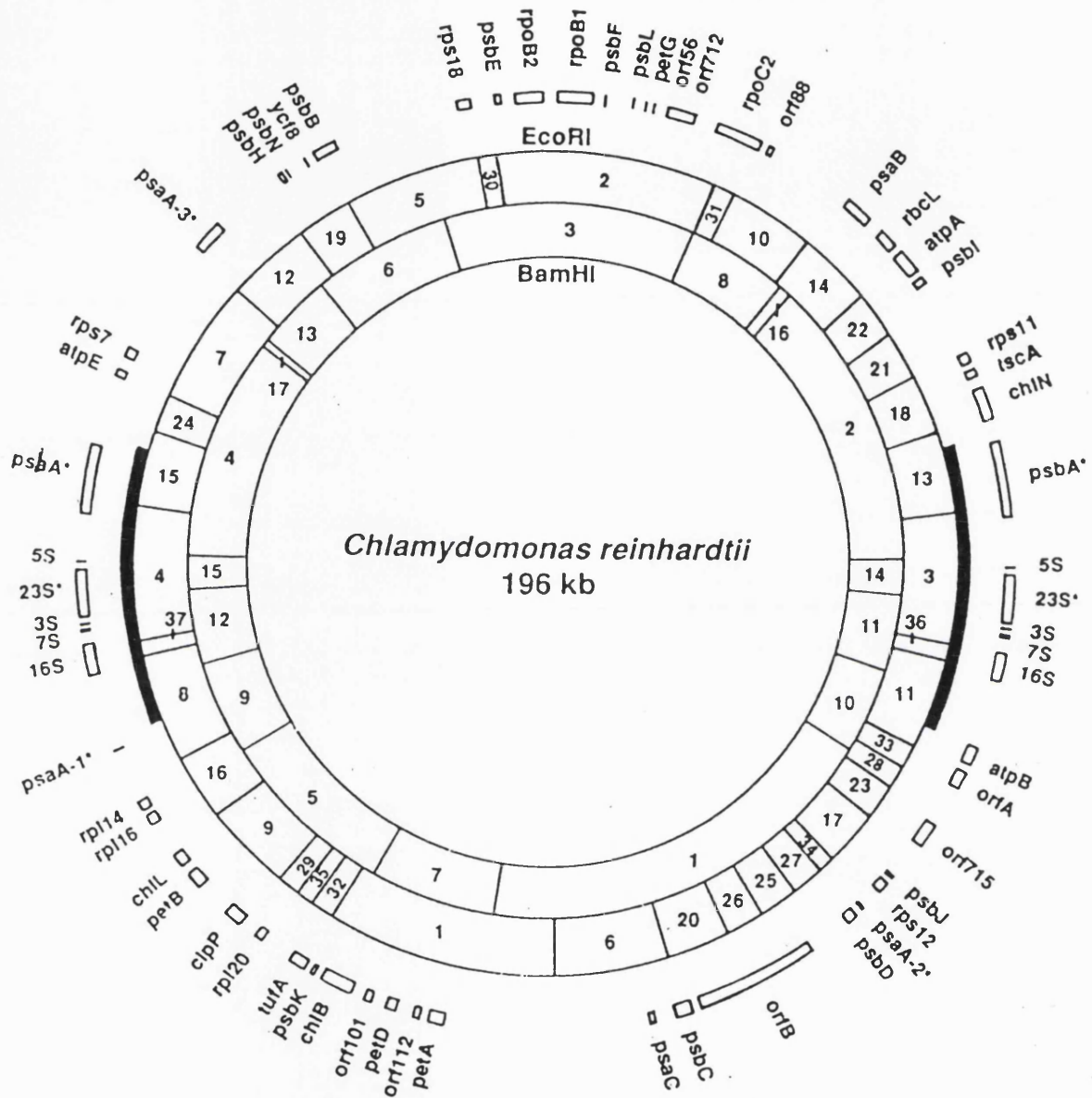
Figure 1.3 *The chloroplast genome of* Chlamydomonas reinhardtii. *The inner and outer circles represent the* Bam*HI and* Eco*RI restriction maps respectively. The thick lines represent inverted repeat regions. Sequenced genes are represented as boxes (tRNA genes not shown). Split genes are marked with an asterisk. Courtesy of Dr. Saul Purton.*
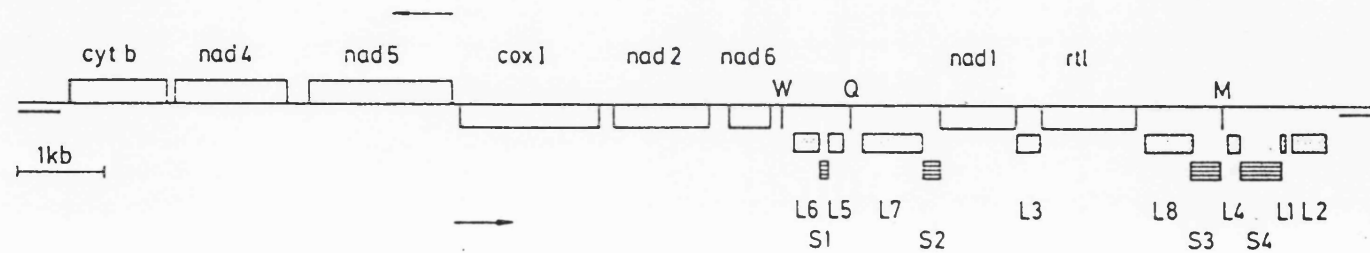
Figure 1.4 *Functional map of the 15.8 kb mitochondrial genome of* Chlamydomonas reinhardtii. *Genes for proteins:* cyt b *encodes apocytochrome* b, nad1, nad2, nad4, nad5 *and* nad6 *encode subunits of NADH dehydrogenase,* cox1 *encodes subunit 1 of cytochrome oxidase and* rt1 *encodes a reverse transcriptase-like protein. The letters* W *(tryptophan),* Q *(glutamine) and* M *(methionine) denote tRNA genes. Regions of the large* (L) *and small* (S) *scrambled rRNA genes are indicated. The inverted repeats at both linear termini are characterised by a double line. Directions of transcription are shown by arrows.*

22

both dominant and recessive mutations can be scored (figure 5.5). Recessive mutations show a phenotype in haploids unlike diploid or polyploid genomes where such mutations would be masked by the wild-type allele. The genetics of *Chlamydomonas* are well defined, and it is amenable to analysis by standard genetic, molecular and biochemical methods. Haploid cells exist as either mating-type (+) or mating-type (-) and can be propagated vegatively. Upon starvation of nitrogen and blue light illumination vegative cells develop into gametes. Mixing of gametes of opposite mating types leads to their fusion and the resulting zygote can be induced under appropriate environmental conditions to undergo meiosis and produce a tetrad consisting of four haploid progeny cells. Genetic analysis of mutants is straightforward; after mating, the four meiotic progeny are examined for Mendelian (2:2) inheritance of the defect and linkage to other markers. Diploid cells can also be generated artificially by somatic fusion and are a useful means of testing for complementation. Mutations in each of the genetic systems can readily be distinguished in crosses, because nuclear genes segregate according to Mendelian rules, whereas chloroplast and mitochondrial genomes are usually transmitted uniparentally to the progeny from the mating type (+) and mating type (-) respectively.

The relatively recent development of efficient methods for transformation of the nuclear, chloroplast and mitochondrial DNA has greatly increased the value of this photosynthetic alga and is discussed in detail later in the chapter. The ability of *Chlamydomonas* to perform functions that other popular model organisms such as yeast do not (i.e. photosynthesis, motility and response to light) make it the system of choice for study of these phenomena.

The photosynthetic apparatus of *Chlamydomonas* is quite similar to that of higher plants (figure 1.6, 1.7). Work on *Chlorella, Scenedemus* and *Chlamydomonas* has been of great importance for understanding the photosynthetic machinery. Among these algae, *Chlamydomonas* offers a special advantage because it lends itself to tetrad analysis and because its photosynthetic function is dispensable when the cells are grown in the presence of a reduced carbon source such as acetate (Levine 1968). This property has allowed for the isolation of numerous mutants that have been invaluable in examining photosynthetic function.

Chlamydomonas is one of the principle eukaryotes used for the genetic and biochemical analysis of flagellar assembly and motility (Johnson 1995). Its two flagella have the typical '9+2' axoneme structure of eukaryotic flagella and cilia. Diverse flagellar mutants can be isolated with ease and many such mutants with defects in flagellar assembly and function have been isolated including those with flagella of aberrant size, number or those that are paralysed or do not beat correctly due to faulty assembly of the axoneme. The individual polypeptides defective in some of these mutants have been identified and their genes have been cloned. Furthermore, the ability of the organism to dispense with photosynthesis and motility avoids the need for conditional mutations or for maintenance of heterozygous
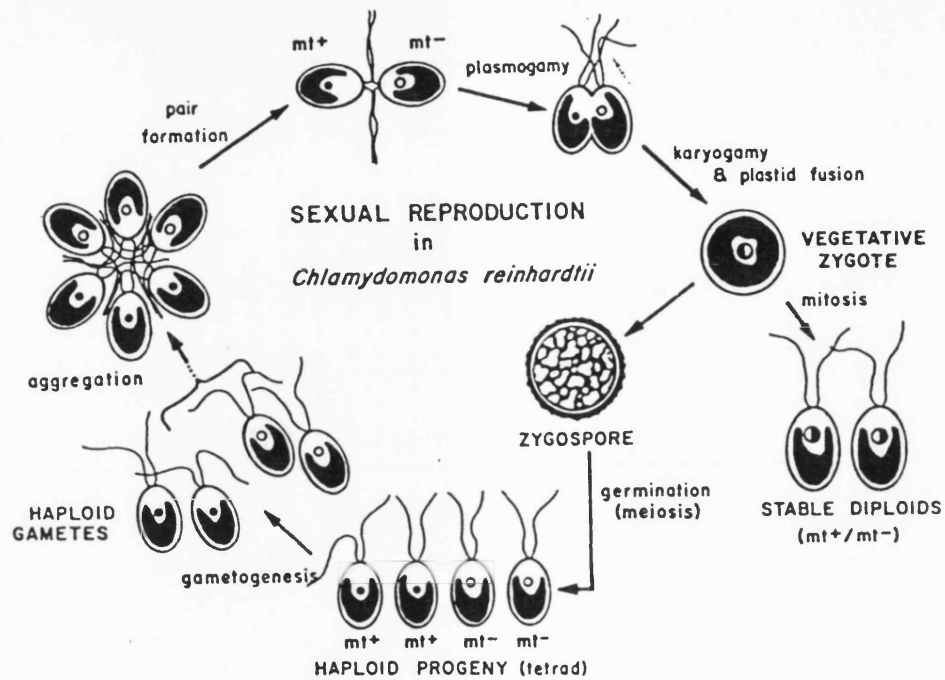
Figure 1.5 *Life cycle of* Chlamydomonas reinhardtii, *showing alternative fates of mated pairs as meiotic zygotes and as vegetative diploid cells. From a drawing by Karen VanWinkle-Swift.*

24
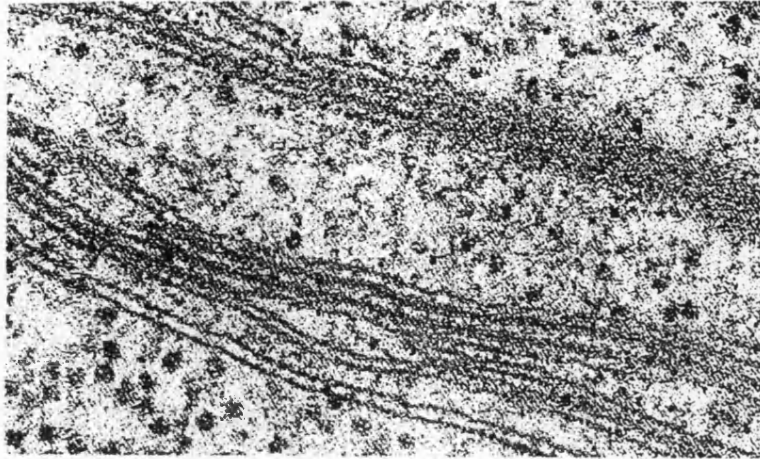
Figure 1.6 *Electron micrograph of* Chlamydomonas reinhardtii *chloroplast, showing the chloroplast envelope, stacked and unstacked thylakoid membranes and cytoplasmic and chloroplast ribosomes. From Bourque et al. 1971.*
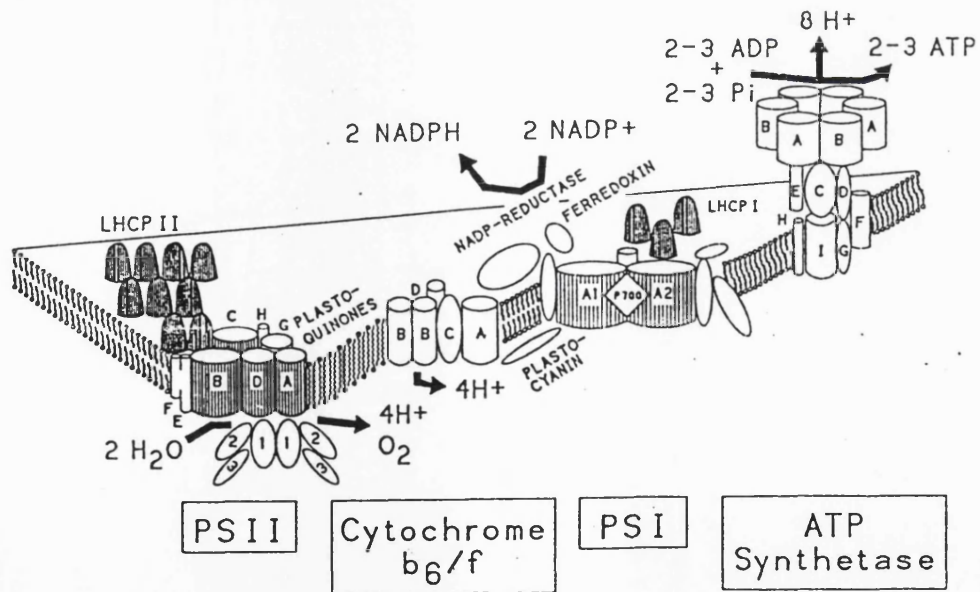
Figure 1.7 *Schematic view of thylakoid membrane components involved in photosynthetic electron transfer. Nuclear-encoded components are represented by ovals and chloroplast-encoded components by cylinders. From a drawing by Gregory Schmidt.*

25

diploid cells. Another feature of *Chlamydomonas* is that it has a primitive eyespot with rhodopsin as the photoreceptor (Sagar and Palade 1967). There is considerable interest in the mechanism of light perception and the signal transduction pathways connecting perception to response (Witman 1993). Mitochondrial mutants of *Chlamydomonas* that lack mitochondrial DNA and die in the dark have been isolated (Randolph-Anderson, Boynton et al. 1993). Their existence has revealed that respiratory function is dispensable in *Chlamydomonas* provided photosynthetic function is intact, paving the way for studies of chloroplast/mitochondrial interaction.

There are numerous resources available to scientists throughout the world who wish to use *Chlamydomonas* in their studies. A large collection of *Chlamydomonas* mutants is maintained at the Chlamydomonas Genetics Centre at Duke University (Harris 1984). The World Wide Web *Chlamydomonas* Resource Page, an Internet Discussion Group, the existence of a map of chloroplast DNA and the large and rapidly growing Genetic, RFLP and physical maps of the nuclear genome greatly facilitate the use of the organism as a model system.

1.1.5 The transformation of *Chlamydomonas*

In vitro alteration of the primary sequence of a gene and reintroduction of the mutated DNA into the genome of a living cell allows direct manipulation of protein structure and has been widely used in exploring biochemical and developmental pathways. The development of recombinant DNA technology for organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Dictyostelium discoideum* and *Arabidopsis thaliana* has greatly increased the scope of investigation that can be undertaken with these organisms. DNA encoding key enzymes, regulatory proteins or structural proteins can now be isolated, modified and reintroduced into an organism so that classical genetic techniques can be combined with modern techniques for DNA cloning and transformation. For example, modification of gene activity may be achieved by over-expression by introducing additional copies into the genome or down-regulation by the application of antisense technology. Endogenous genes can also be manipulated via gene disruption (gene knockout) or site directed mutagenesis to produce new mutant phenotypes. Methods for the generation of 'tagged' mutants as a result of the integration of transforming DNA into the genome and complementation studies also become possible. Similarly, the expression of foreign genes to confer a novel phenotype or for production of a recombinant protein. The fact that prokaryotic and eukaryotic genes have differing structures in terms of non-translated controlling elements and presence of introns means that where heterologous DNA is eukaryotic in origin it is often desirable to express the DNA in a eukaryotic host rather than a prokaryote in order that the gene may be properly processed.

26

An obvious prerequisite for genetic engineering is the ability to introduce DNA into a living cell in such a way that it is stably maintained and properly expressed in the host cell. It is possible to transform the chloroplast, the nuclear and the mitochondrial DNA of *Chlamydomonas*. The biolistic process, in which DNA-coated microprojectiles are propelled at high velocity into eukaryotic cells under partial vacuum, was first used for stable transformation of the nuclear genome of plant cells (Klein, Wolf et al. 1987). Biolistic transformation of all three genomes of *Chlamydomonas* is possible-the nuclear (Boynton, Gillham et al. 1989) the chloroplast (Boynton 1988) and the mitochondrial (Randolph-Anderson, Boynton et al. 1993). One element of any successful transformation strategy is selection i.e. how to separate transformed cells from non-transformed cells. The most powerful strategy is that which in transformed cells gain the ability to grow on media on which non-transformed cells die. The first successful transformations were selected by mutant rescue i.e. restoration of wild-type phenotype to mutant cells by introduction of wild-type DNA into those cells.

In 1988, Boynton et al. reported transformation of the chloroplast DNA of
(Boynton, Gillham et al. 1988)
Chlamydomonas using biolistics. Selection for transformants was achieved through restoration to photosynthetic growth of non-photosynthetic host cells. Mutants defective in the chloroplast *atpB* gene were bombarded with tungsten particles carrying the cloned wild-type gene encoding the missing product. Since then much progress has been made. Several markers are now available for selection or as reporters for the chloroplast (Goldschmidt-Clermont 1991; Boynton and Gillham 1993; Erickson 1996) including positive selective markers that confer resistance to antibiotics (Harris 1989; Goldschmidt Clermont 1991), herbicides and phytotoxins (Erickson 1996). Integration of DNA into the chloroplast occurs through homologous recombination into one or more of the resident chloroplast DNA molecules. The integrated DNA replaces the homologous segment of the host DNA molecule. Non-homologous DNA can be introduced into the chloroplast genome provided it is flanked by DNA with homology to the chloroplast DNA. Chloroplast gene disruptions can be either heteroplasmic or homoplasmic. Initially a transformed cell will be heteroplasmic because its population of chloroplast DNA molecules is no longer homogenous but consists of resident host DNA molecules plus newly formed recombinant DNA molecules. Where the gene disrupted is not essential to growth, such as in the case of a photosynthetic protein, homoplasticity may be attained (i.e. all the DNA molecules are the same) by maintaining the selective pressure. However, where an essential gene has been interrupted, heteroplasticity must be retained in order that the cell will remain viable.

Stable nuclear transformation of *Chlamydomonas* was achieved in 1989 with the use of biolistics to rescue arginine-requiring mutants (*arg7*) to arginine auxotrophy by shooting them with DNA encoding the wild-type gene, *ARG7* (Debuchy, Purton et al. 1989).

27

Basis of Selection	Host Strain	Selectable Marker	Selective medium and conditions
1) Non-photosynthetic host; chloroplast gene as selectable marker	Deletion in chloroplast gene:		Minimal medium, incubation in light
	atpB	WT *atpB*	
	psbA	WT *psbA*	
	tscA	WT *tscA*	
	Chloroplast point mutant:		
	atpB (transversion)	WT *atpB*	
	psaB (frameshift)	WT *psaB*	
2) Chemical resistance marker gene:			Acetate medium plus selective chemical, light or dark incubation
a) *Chlamydomonas* chloroplast gene as selectable marker	Ribosome as target for antibiotics:		
	16S rRNA	streptomycin/ spectinomycin resistance	
	23S rRNA	erythromycin resistance	
	PSII Q_b site as target for herbicides:		
	psbA, codon 264	metribuzin DCMU	
	Proton ATP synthase CF1 as target for phytotoxins:		
	atpB, codon 83	tenoxin	
b) Bacterial *aadA* as selectable antibiotic resistance marker	Any	streptomycin/ spectinomycin resistance	

Table 1 b *Chloroplast transformation markers*

Subsequently, other mutant phenotypes were rescued by this approach. For example, a non-photosynthetic mutant lacking the gene encoding oxygen evolving complex 1 was restored to full auxotrophy by shooting in the wild-type gene for *oee1* (Mayfield and Kindle 1990a). Similarly, mutants lacking in nitrate reductase activity have been restored to full auxotrophy by transforming with the gene encoding nitrate reductase (Fernandez, Schnell et al. 1989). However, transformation rates were low and it was found that the transformants harboured multiple copies of the transforming DNA. Improvement in the transformation efficiency came with the introduction of the glass bead transformation system for the nuclear genome of *Chlamydomonas* (Kindle 1990). This method has proved rapid, reliable and easy to use. Consistently high levels of transformation of the nuclear genome (> than 10^{-5} - equivalent to ~ 500 colonies per plate) can now be achieved by agitating cell wall-deficient mutants of *Chlamydomonas* with glass beads and the transforming DNA. Where it is desirable to transform a host that has a cell wall, the cell wall can be transiently removed by treating with autolysin (Kindle 1990). The DNA enters the cell where it integrates into the genome by non-homologous recombination i.e. at random positions. This method usually results in only a few copies of DNA in the transformant genome. The presence of only one copy in most transformants can be achieved by limiting the concentration of DNA. *Chlamydomonas* has subsequently been transformed by other methods. Electroporation has been used to introduce plasmids up to 14 kb in size into cells with or without cell walls although transformation rates were low (only 17 colonies from 10^7 treated cells compared with 10^4 cells using the glass bead transformation method) (Brown, Sprecher et al. 1991). More recently, high transformation efficiencies have been reported using electroporation (Tang, Qiao et al. 1995). Silicon carbide whiskers have also been used to transform walled cells at rates comparable to the glass bead transformation method (Dunahay 1993).

Mitochondrial transformation of *Chlamydomonas* has been demonstrated, but has yet to be exploited (Randolph-Anderson, Boynton et al. 1993). A respiratory deficient mutant *dum*-1 that fails to grow in the dark due to a terminal 1.5 kb deletion in the mitochondrial genome was restored to full respiratory activity by biolistic bombardment with *Chlamydomonas* mitochondrial DNA.

The development of these transformation systems has heralded an avalanche of studies in our understanding of *Chlamydomonas* genes and how they function. More recently, new dominant selectable markers that confer resistance to drugs have been developed therefore allowing any strain of *Chlamydomonas* to be used. The next section of this introduction offers a précis of the achievements so far in the field of nuclear transformation, as well as covering important advances in the field of chloroplast transformation, as the two are by no means mutually exclusive.

1.2 Molecular techniques in *Chlamydomonas*

1.2.1 Chloroplast markers and heterologous gene expression in the chloroplast

A variety of selectable markers have been developed for use in the chloroplast. Mutant rescue of non-photosynthetic hosts to full auxotrophy using *Chlamydomonas* chloroplast genes has been used as well as selection of transformants using endogenous chloroplast genes that encode resistance to antibiotics, herbicides or chemical toxins. These include streptomycin, spectinomycin and erythromycin resistance encoded by ribosomal DNA, herbicide resistance encoded by the *psbA* gene (Boynton and Gillham 1993) and a trans-splicing function provided by the *tscA* gene (Goldschmidt Clermont 1991). The *aadA* cassette, which utilises a bacterial gene encoding aminoglycoside resistance is an example of a transgenic selectable marker for the chloroplast and heterologous gene expression has also been achieved with the GUS reporter system, encoded by the bacterial *uidA* gene (Kindle, Richards et al. 1991).

It has been difficult to obtain chloroplast expression of foreign genes at the level of protein product. Lack of information on the chloroplast promoter and the effect that 5' to 3' untranslated regions may have on message stability and translation hindered many early attempts on chloroplast transgenic expression. Success was achieved in 1991 using the bacterial *aadA* gene encoding aminoglycoside adenyltransferase, which confers resistance to spectinomycin and streptomycin (Goldschmidt Clermont 1991). The bacterial gene for *aadA* was placed under control of the 3' and 5' sequences from various chloroplast genes in a series of expression vectors and transformed into the chloroplast genome. These upstream and downstream sequences allowed the non-homologous *aadA* DNA to integrate into the genome through homologous recombination with the corresponding sequences. Such targeting allows for placement of the dominant selectable marker in a neutral site in the chloroplast genome or for targeting of the marker to a specific gene, thus *aadA* is a powerful tool as a selectable marker and for creating insertion and replacement mutations.

The bacterial *uidA* gene, encoding β-glucoronidase (GUS), has been used extensively as a reporter gene in plant and agricultural molecular biology (Jefferson 1989). β-glucoronidase can cleave a wide variety of artificial chromogenic substrates, providing a colourimetric assay for gene expression. Chloroplast transformants of *Chlamydomonas* containing transcriptional fusions between a *Chlamydomonas* chloroplast promoter region and the bacterial gene showed transcripts for the foreign gene but no detectable gene product. Subsequent investigations showed that inclusion of the 5' untranslated region of the chloroplast *petD* gene stabilise the transcript and the transformants accumulated a high level of GUS activity. These translational fusions extended the N-terminus by 9 amino acids and appeared to facilitate translation of the chimeric mRNA (Sakamoto, Kindle et al. 1993).

Expression of bacterial *NPTII*, encoding the bacterial neomycin phosphotransferase structural gene, has so far been attained only at the transcriptional level in the chloroplast (Blowers, Bogorad et al. 1989).

Co-transformation of markers on independent plasmids is efficient and it is therefore possible to introduce a gene with an unselected phenotype and identify transformants by virtue of a co-transformed selectable marker. By co-transforming a selectable marker along with an appropriately altered chloroplast DNA, chloroplast genes may be mutated or eliminated from the genome. Homologous recombination in the chloroplast has been exploited for targeted insertional inactivation and mutagenesis of *Chlamydomonas* chloroplast genes. A gene is disrupted or modified *in vitro* and subsequently reintroduced to the host. For insertional inactivation of chloroplast genes, any DNA may be used to disrupt a particular gene and the DNA should be flanked by chloroplast DNA encoding the gene of interest. For example, yeast DNA has been used as an insertional mutagen (Newman, Gillham et al. 1991). A construct carrying a chloroplast gene disrupted with 0.48 kb of yeast DNA was introduced by co-transformation along with a second construct encoding a marker gene. The transformants were subsequently screened for the presence of the yeast DNA. However, the use of a heterologous marker molecule such as *aadA* as an insertional mutagen avoids the need for co-transformation since *aadA* can be used for direct screening of the transformants.

In the relatively short time since the advent of chloroplast transformation, there have been an impressive number of reports exploiting the new technology to investigate many facets of chloroplast molecular biology. These studies include the identification of novel genes through the disruption of open reading frames: the use of reporter genes to study expression levels, and the role of nuclear gene products in chloroplast gene expression, and structure-function studies of the photosynthetic apparatus through site-directed mutagenesis of the chloroplast encoded subunits.

1.2.2 Nuclear markers and heterologous gene expression in the nuclear genome

Strains with defective nuclear genes can be used as recipients for transformation. Selection is performed by transforming the mutant with the corresponding cloned wild-type gene, and restoration of the wild-type phenotype. Several *Chlamydomonas* mutants have been used including those lacking in nitrate reductase activity (*NIT1*) (Fernandez, Schnell et al. 1989; Kindle 1990), argininosuccinate lyase activity (*ARG7*) (Debuchy, Purton et al. 1989) and oxygen evolving enhancer protein 1 (*OEE1*) (Mayfield and Kindle 1990). The gene coding for radial spoke protein 3 (*RSP3*) has also been used for rescue of a paralysed-flagella mutant and may be used for selection of transformants (Diener, Curry et al. 1990). Kindle demonstrated that mutant rescue can be used to indicate introduction of a separate non-

31

selected gene by co-transformation. The second gene is introduced as a separate molecule along with the marker and the transformants subsequently screened for the presence of the second transforming DNA. It was found that up to 60% of transformants isolated by this method are co-transformants allowing for the introduction of any type of DNA into the genome (Kindle 1990).

More recently, dominant selectable markers have been developed that confer resistance to drugs and thereby allow non-mutant strains of *Chlamydomonas* to be transformed. The cloned mutant gene, *CRY-1-1*, confers cryptopleurine and emetine resistance when transformed into wild-type cells. When fused to the strong, constitutive promoter of the small subunit of ribulose-bisphosphate carboxylase/oxygenase (*RBCS2*), the level of resistance can be substantially increased. This increased expression of *CRY1* afforded by the *RBCS2* promoter allowed direct selection of transformants on emetine and can be used to introduce other types of DNA into the genome by co-transformation (Nelson, Savereide et al. 1994).

A milestone in the development of further dominant selectable markers for *Chlamydomonas* has come with the ability to stably introduce and express heterologous DNA in the nuclear genome. Until recently it has not been possible to express foreign genes in the nucleus of *Chlamydomonas*. A few reports of selection using heterologous genes have appeared (Rochaix and Dillewijn 1982; Hasnain, Manavathu et al. 1985; Hall, Taylor et al. 1993), but these results were not found to be reproducible and these heterologous genes did not allow efficient recovery of transformants. A variety of reasons have been suggested for the poor expression levels, including silencing of introduced foreign genes by methylation, differences in codon usage, ectopic effects and lack of suitable coding elements such as promoters and introns. Recently, the coding sequence of a gene conferring phleomycin resistance (*BLE*) from *Streptoalloteichus hindustanus* has successfully been expressed and used for selection of transformants. This gene had been chosen due to the similarity in its codon usage to that of nuclear genes of *Chlamydomonas* and its short coding sequence of 387 bp. A chimeric gene was made in which *BLE* was fused to the 3' to 5' untranslated regions of *RBCS2* and used for transformation (Stevens, Purton et al. 1996). The transformants could be selected directly on drug-containing medium provided they were allowed to recover for 18 hours prior to selection. Increased levels of expression of *BLE* have been achieved by truncating the promoter region and inclusion of an intronic enhancer element in the 5' UTR of *RBCS2* (Lumbreras, Stevens et al. 1998).

The aminoglycoside 3'-phosphotransferase typeVIII encoding gene (*aphVIII*) from *Streptomyces rimosus* has also been used for direct selection of transformants of *Chlamydomonas* (Sizova 1996). *AphVIII* was again chosen for its codon usage. The *aphVIII* structural gene flanked by *S. rimosus* regulatory sequence was transformed into *Chlamydomonas*. The integrated *Streptomyces* DNA is stably maintained and the

transformants show APHVIII enzyme activity. It seems likely from the frequency of transformation to paramomycin resistance (1.3 - 1.9 x 10^7) that expression occurs as a result of fusion with *Chlamydomonas* regulatory elements within the nuclear genome. Other reports of heterologous DNA expression in *Chlamydomonas* include expression of chloramphenicol acetyltransferase (CAT) gene which had been introduced to the genome by electroporation. Expression of *cat* was confirmed by detection of a *cat* transcript in northern analysis and the detection of CAT enzyme by the ELISA assay (Tang, Qiao et al. 1995). Expression of the eubacterial *aadA* gene fused to the 5' and 3' untranslated regions of the endogenous *RBCS2* gene in the nucleus has been achieved (Cerutti, Johnson et al. 1997). *AadA* confers spectinomycin resistance and has also been used in the chloroplast (see section 1.2.1). Spectinomycin resistant transformants have been isolated by direct selection and contain the chimeric gene stably integrated into the nuclear genome in genetically linked multiple copies. They also show co-segregation of the resistance phenotype with the introduced DNA, and synthesis of the expected mRNA and protein.

The demonstration that heterologous DNA can be stably maintained and expressed in the *Chlamydomonas* nuclear genome represents a significant leap forward in the understanding of the factors required for the efficient expression of heterologous and endogenous DNA in the nuclear genome of *Chlamydomonas*.

1.2.3 Insertional mutagenesis

The apparently random integration of transforming DNA into the nuclear genome of *Chlamydomonas* has been exploited by many groups as a method for insertional mutagenesis. Random integration of DNA can disrupt a functionally important gene, creating a 'tagged' mutant. One of the earliest successes was the use of the wild-type nitrate reductase gene to transform *nit1ˉ* mutants which produced over 3000 cells carrying random inserts of the *nit1* gene (Tam and Lefebvre 1995). When these transformants were screened for defects in mobility, 74 potential mutants were obtained, the phenotypes of which were quite variable. It was found that the mutant phenotype co-segregates with the newly acquired nitrate reductase gene, and that such tagged mutants are stable and amenable to molecular analysis as the introduced DNA can be used to isolate the affected gene. In tests with two of the mutants, plasmid rescue of sequences from the interrupted gene was made possible by the fact that sequences from the parent *E. coli* plasmid had integrated along with the nitrate reductase gene. These flanking sequences were then used to isolate the wild-type copy of the interrupted gene from a genomic library. Transformation with the wild-type gene resulted in the restoration of the wild-type phenotype in both of mutant phenotypes.

Using this approach, a wide range of mutant phenotypes have been found, including those affected in photosynthesis (Adam, Lentz et al. 1993; Gumpel, Ralley et al. 1995), respiration (Turner, Gumpel et al. 1996) sulphur metabolism (Davies, Weeks et al. 1992),

phototaxis (Parzour, Sineshchekov et al. *1995*), cadmium tolerance (McHugh and Spanier 1994) and salt tolerance (Prieto, Pardo et al. 1996). A variation of the gene tagging technique is the promoter trap system (Haring and Beck 1997). The system involves the transformation of an *arg7 pf14* double mutant in which *pf14* is a *paralysed flagella* mutation caused by a defect in *RSP3* (radial spoke protein 3). Transformation is achieved using the *ARG7* marker (for selection of transformants) and a promoterless version of *RSP3* (the promoter trap). Random integration into the nuclear genome occasionally results in insertion of *RSP3* downstream of an endogenous promoter, thereby generating motile transformants. The endogenous gene whose promoter is driving *RSP3* expression can be cloned by virtue of its linkage to the *RSP3* DNA. Tagged mutants are identified by their motility, and this system can be used to isolate genes that are conditionally expressed.

1.2.4 Transposon tagging

In higher plants and other eukaryotes such as *Drosophila* transposon tagging has been exploited as a means of generating tagged mutants. Transposons are mobile DNA elements that can move around in the genome and are often associated with the appearance of new mutations (Walbot 1992). This property has important evolutionary consequences and has been important for the isolation of transposons from a wide range of species.

A variety of transposable elements have been identified in *Chlamydomonas,* some of which appear to be mobile enough to warrant attention as potential gene tagging agents. The *NIT2* gene, a component of the nitrate-assimilation pathway was isolated using a transposon tagging strategy (Schnell and Lefebvre 1993). A *nit2* mutation caused by the insertion of a transposon was identified by testing spontaneous *nit2⁻* mutants for the presence of *Gulliver* or *TOC1* (Day and Rochaix 1991) transposable elements that have been identified in *Chlamydomonas.* In two of 14 different mutants that were analysed, a *Gulliver* element was found to be genetically and phenotypically associated with the *nit2⁻* mutation. Using the *Gulliver* element as a probe, one of the transposon-induced *nit2* alleles was isolated. A sequence adjoining the transposon was then used to isolate the corresponding wild-type locus. Using the cloned *nit2* gene as a probe they then surveyed eight of the remaining spontaneous mutants with apparent inserts or rearrangements in *nit2* and found two additional transposable elements. Thus, the list of active transposable elements that could serve as important tools for tagging and recovering spontaneous recessive mutants in *Chlamydomonas* continues to grow.

1.2.5 Genomic complementation

The ability to complement a given mutant with a cloned wild-type gene allows the correlation of genetic and molecular data. As previously mentioned, numerous nuclear mutants of *Chlamydomonas* have been found that are affected in many important cellular

functions (Harris 1984). In most cases the genes responsible for these mutations are still unknown. Nuclear mutants can be rescued successfully by complementation with a pooled wild-type cosmid library (Purton and Rochaix 1994). The mutant complemented was a cell-wall deficient strain that carries a defect in the gene encoding argininosuccinate lyase and is unable to grow in the absence of exogenous arginine. Total DNA from the cosmid library was delivered to approximately 2×10^9 cells via the glass bead transformation system. Four putative transformants were recovered, three of which were demonstrated to have acquired an extra copy of the gene *ARG7*, showed a stable phenotype and high levels of ASL activity.

To improve the efficiency with which genes complementing a particular mutation can be isolated an indexed library of 11,280 ordered clones has been established in microtitre plates (Zhang, Herman et al. 1994). Also, rapid screening methods with cosmid DNAs pooled from individual microtitre dishes have been applied successfully. This allows the rapid isolation of genes using subsets of the library to pinpoint the clone that complements a given mutant (Funke, Kovar et al. 1997). More recently a *Chlamydomonas* genomic library has been constructed using yeast artificial chromosomes (YACs). It has been demonstrated that clones from a genomic YAC library could also be used in the isolation of genes by complementation (Vashishtha, Segil et al. 1996).

1.2.6 Homologous recombination

It is desirable to generate a mutation in a cloned gene and to substitute the mutated gene for the endogenous gene by homologous recombination; an approach that has been widely exploited in bacteria and lower eukaryotes. In *Chlamydomonas* the development of efficient methods for gene targeting has been hindered by the low frequency of homologous recombination in the nuclear genome of *Chalmydomonas* (Debuchy, Purton et al. 1989; Kindle, Schnell et al. 1989). The only convincing report of targeted disruption of a nuclear gene in *Chlamydomonas* involved the disruption of *NIT8*, which encodes a protein necessary for nitrate and nitrite assimilation, and required a second selective marker to identify the few *CRY1-1* transformants affected in the assimilation pathway (Nelson and Lefebvre 1995). A vector was constructed in which *NIT8* was disrupted by placing the *CRY1-1* selectable marker within the *NIT8* coding region. *Nit8⁻* mutants were selected on emetine (positive selection for *CRY1-1* transformants) and chlorate (negative selection against any mutants not affected in nitrate assimilation). One of 2000 transformants obtained after selection on emetine/chlorate medium was found to contain a homologous insert of five copies of the disruption plasmid into the *NIT8* gene. The gene had been successfully disrupted by targeting, although the frequency of the event was extremely low, even using direct selection for the desired mutant. The mutant phenotype was rescued by transformation

with the wild-type *NIT8* gene, confirming the *NIT8* mutation was the sole genetic lesion responsible for the nitrate phenotype.

Other groups have achieved the same goal of gene knockout by screening large populations of transformants for those in which the target gene has been interrupted by random insertion of the transforming DNA (Wilkerson, King et al. 1995; Khrebtukova and Spreizer 1996; Myster, Knott et al. 1997). A general and routine system for altering the nuclear genotype of *Chlamydomonas* through gene targeting is an important goal in the development of the model system.

1.2.7 Reporter molecules

Although attempts to use heterologous genes encoding reporter molecules such as β-glucoronidase or luciferase have so far proved disappointing, two *Chlamydomonas* genes have been employed as reporters of promoter activity. The gene for aryl-sulphatase (*ARS*) has proved to be useful as such a reporter (Ohresser, Matagne et al. 1997). This gene encodes an enzyme produced by *Chlamydomonas* under conditions of sulphur starvation which is secreted into the periplasm and can be readily assayed using the chromogenic substrate 5-bromo-4-chloro-3-indolylsulphate (X-SO₄). The coding sequence of *ARS* has been linked to promoters from different genes to form chimeras. The activity of the promoter can then be monitored for assaying for cleavage of the chromogenic substrate X-SO₄ (Davies, Weeks et al. 1992). Davies and Grossmann$_\wedge$were$^{(1992)}$ able to use this approach to identify *cis*-acting elements within the promoter region of the β-tubulin gene *TUBB2* that were required for the induction of *TUBB2* transcription following deflagellation of *Chlamydomonas* cells. Similarly, the *RSP3* gene, encoding radial spoke protein can be used as a reporter of promoter activity when introduced into the *pf14* paralysed flagella mutant background. In this case restored cell motility is the assay of promoter activity (Zhang, Herman et al. 1994).

1.2.8 Epitope tagging

Protein localisation studies are possible by epitope tagging of cloned genes, and epitope tagging has been used to study location and accumulation of recombinant protein in transformed cells of *Chlamydomonas* (Kozminski, Diener et al. 1993). Acylation of the flagellar gene for α-tubulin has been shown to occur in *Chlamydomonas* and in a variety of
   | related
evolutionari$_\wedge$organisms, although the function of this acylation remains unclear. In order to study the function an acetylated isoform of α-tubulin in *Chlamydomonas*, the organism was transformed with an α 1-tubulin whose product cannot be acetylated by replacement of the acetylable lysine-40 with codons for non-acetylable amino acids. The mutagenised

36

tubulin was tagged with an epitope from the influenza virus haemagluttinin (HA). Using antibodies against the HA epitope, the presence of the non-acetylable variant in both cytoplasmic and flagellar microtubules could be investigated. Although clones with high level expression of non-acetylable α-tubulin expression were observed using this approach, no gross phenotypic effect was observed.

A second, simpler approach to epitope tagging has recently been demonstrated by Jarvik et al. (Jarvik, Adler et al. 1996). CD-tagging puts specific molecular tags on a gene, its transcript and its protein product. The method uses a specially designed DNA molecule, the CD-cassette, that contains splice acceptor and donor sites surrounding a short open-reading frame of 18 amino acids that comprises the epitope. By introducing this cassette into the intron of a cloned gene, a 'mini-exon' is created, bordered by two hybrid introns. Expression of the gene in *Chlamydomonas* results in a protein containing an additional epitope sequence. The mRNA is also tagged with a specific nucleotide sequence, as is the gene itself. As these tags are unique, specific nucleotide or antibody probes can be used to obtain/analyse the gene, transcript or protein. Jarvik used this approach in tagging the *RSP3* gene of *Chlamydomonas*, a three intron gene that encodes the flagellar radial spoke protein, RSP3. The cassette was introduced into an intron of *RSP3*, and it was shown that the gene was still able to rescue a non-motile *pf14* mutant and that antibodies against the epitope were able to detect the protein in the flagellar axonemes.

1.2.9 Prospects for molecular techniques in *Chlamydomonas*

Chlamydomonas has already become established as a powerful model system for the molecular-genetic studies of cellular processes, particularly motility and photosynthesis. The isolation of genes by genomic complementation or gene tagging, combined with the generation of stable mutants by insertional mutagenesis, provides the tools for detailed study of gene expression and protein function. The molecular and genetic maps of the *Chlamydomonas* nuclear genome are constantly expanding and the sequencing of the chloroplast genome is almost complete. It is anticipated that other molecular tools and techniques will become available in the near future. Desirable prospects include an increased repertoire of reporter molecules, dominant markers and epitopes for gene tagging and the development of negative selectable markers, resulting in a routine procedure for gene targeting and enrichment for rare homologous transformation events. Finally manipulation of gene expression levels should be possible using RNA antisense to down-regulate target genes or strong promoters to over-express these genes.

The next section of this introduction details the use of Expressed Sequence Tag analysis, Green Fluorescent Protein as a reporter molecule and the possibility of developing a

negative selectable marker for homologous transformation in *Chlamydomonas*. Antisense technology as a means of downregulation of specific genes is also covered.

1.3 Analysis of Expressed Sequence Tags (ESTs)

1.3.1 The use of ESTs in genome sequencing projects

Systematic genomic and cDNA sequencing projects are currently underway for several organisms. Despite the development of rapid, automated sequencing methods, the establishment of the complete sequence for even the smallest of eukaryotic genomes can take many years. Several model organism genomes have now been extensively sequenced including the successful completion of the *S. cerevisiae* sequencing project (Goffeau and al 1996). Insights from model organisms provide valuable clues to gene function for higher organisms and are an integrated part of the Human Genome Project (Strachan, Abitol et al. 1997). The first stage in building up a complete sequence map involves the construction of a genetic linkage map. Genetic linkage maps consist of DNA markers, spaced over the entire genome. The markers usually comprise DNA polymorphisms such as mutations in phenotype, drug resistance markers or polymorphic repeat sequences such as microsatellites. These markers are then organised into linkage groups, which are used as a framework for the integration of the genetic markers into the physical contig map. Physical contig maps are based on the fragmentation of the genome and the cloning of overlapping fragments into vectors for propagation and storage. Originally cosmids, capable of taking fragments of genomic DNA as large as 45 kb, were used. More recently, YACs have been employed, capable of accepting inserts as large as 1-2 Mb. YACs have become the standard tool for the construction of chromosome and whole genome contig maps. Other vectors include PACs and P1s (phage artificial chromosomes) and BACs (bacterial artificial chromosomes) which are less prone to rearrangement (Spurr 1995).

cDNA is an important tool for elucidation of the genome structure and reflects the expression mode of genes during developmental stages, under various environmental conditions and in the different tissue types for higher organisms. Automated single read sequencing from one or both ends of cDNA clones to generate ESTs (Expressed Sequence Tags) is an efficient strategy for identifying new genes and characterising the expressed gene content of cells, tissues and species. More than 80% of the genes identified by EST sequencing in humans are new genes of unknown function. Of new human genes tentatively identified by database searches, 66% are more similar to genes from a variety of non human species than to any previously described human gene. EST analysis can also lead to the identification of functional motifs e.g. where a DNA element codes for a protein structure such as a DNA or RNA binding element, thus providing a clue to the function of the gene product.

The Human Genome Project is an international effort to determine the location and structure of all genes active in a number of organisms. The human genome contains an estimated 50 000 to 100 000 genes, of which only a few thousand have been sequenced. Complex genetic and physical maps now exist for the human genome, and theoretically, it should be possible to isolate a majority of the remaining uncharacterised genes from a set of cDNA libraries representing different tissues and developmental stages. Largely due to the Human Genome Initiative and its spin-offs, technological progress in the field of cDNA analysis have far exceeded expectations in the past decade. Automated DNA sequencing, the production of cDNA microarrays where 1000s of DNA molecules can be printed onto a glass slide, and differential screening strategies where cDNAs expressed in different tissue types or under various physiological conditions are isolated, are now possible (Schena 1996). The project is ahead of schedule and is expected to be completed by 2003, if not earlier (Radich 1996). The commercial implications are vast for the pharmaceutical and biotechnology industries, and the ethical issues are complex, spawning mass public debate and legislation over patenting and ownership rights to this wealth of information. The technological advances keep coming, including the opening of a whole new field, that of 'bioinformatics'.

1.3.2 DNA sequence databases

Repositories for information being generated by genome sequencing strategies exist in the form of public-domain databases held by Genbank, the European Laboratory for Molecular Biology (EMBL), the US National Centre for Biotechnology Information (NCBI) and the DNA Data Base of Japan (DDBJ). Private databases also exist, owned for example by Human Genome Sciences and Incyte Pharmaceuticals. Genbank sequence data exists in the form of full-length, carefully proof-read sequences and short, single pass EST sequences held in a separate database, dbEST. Similarly, protein sequence information banks, such as Protein Identification Resources (PIR) and Swiss-Prot also exist in the public domain. Nucleotide sequences may be translated in the six open reading frames and compared with protein sequence databases using the program BLASTX (Altschul, Gish et al. 1990) or using the program FASTA (Pearson and Lipman 1988). On this basis it is often possible to identify ESTs, or to gain an insight into their possible function.

1.3.3 The generation and analysis of ESTs

The first extensive analysis of a population of expressed genes was made for human brain tissue (Adams 1992). Expressed genes captured as cDNAs were partially sequenced and compared against the Genbank database using the BLAST algorithm and for translated amino acid sequences against the PIR database using the FASTA algorithm. About 17% of the sequenced clones coded for already reported gene products including genes reported

from *E. coli*, *S. cerevisiae*, *C. elegans*, *Drosophila* and many plants. Other human genome analysis groups reported clear differences in gene expression among various tissues using random sequencing. Besides human genes, cDNA clones of many other organisms including *C. elegans* (McCombie 1992) mouse (Hoog 1991) rice (Uchimiya et al 1992) and *Arabidopsis* (Hofte 1993) have been subject to EST analysis.

A complete EST database and clone collection for a model organism for which well developed genetic, transformation and molecular techniques are available will aid the functional identification of new gene sequences. If the organism also has a well developed physical map correlated with the genetic map, physical mapping of ESTs by hybridisation to cosmids or YACs can yield candidate phenotypes and genomic clones for use in transformation rescue in a single step. Such a strategy would prove useful in *Chlamydomonas* as a set of molecular markers for genome mapping and for the isolation of novel genes prior to the *in vivo* analysis of their gene products.

1.4 Green fluorescent protein as a reporter for gene expression

1.4.1 The use of GFP in other systems

Many different organisms, ranging from bacteria and fungi to fireflies and fish have the ability to emit light for a variety of functions (Hastings 1996). The Green Fluorescent Proteins (GFPs) are a unique set of proteins involved in the bioluminescence of certain species of jellyfish. GFP was originally isolated from the jellyfish *Aequorea victoria* in 1962 by Shimomura (Shimomura, Johnson et al. 1962). The protein emits a bright green flash (509 nm) when excited by blue light at around 375 nm. In 1992 the cDNA of GFP was cloned from *Aequorea victoria* (Prasher, Eckenrode et al. 1992). The intrinsic fluorescent properties of this protein were revealed when the coding sequence for GFP (*gfp*) was isolated and expressed in *E. coli*, yielding a fluorescent product identical to that of the native *Aequorea* protein (Chalfie, Tu et al. 1994). The demonstration that expression of *gfp* in the absence of other *Aequorea* proteins results in a fluorescent product indicated the possible use of GFP as a marker for gene expression and protein localisation. All that is needed for the use of GFP as a reporter molecule is the DNA sequence coding for the GFP polypeptide, which can be fused to that of any protein whose expression or location is of interest. Under favourable conditions, the expressed GFP domain folds independently of the protein to which it is fused and accompanies it *in vivo* as an attached fluorescent label.

Further biochemical characterisation of GFP has revealed it to be\29.9 kDa protein of 238 amino acid residues that exists as a monomer in solution. Most coloured proteins contain a bound prosthetic group (e.g. chlorophylls in the photosynthetic systems or haem‚which \bound to haemoglobin, gives blood its red colour). GFP is unusual in that its colour is derived from the auto-catalytic oxidation of part of the polypeptide itself. The peptide undergoes cyclisation of amino acids 65-67 within its primary structure, to produce a cyclic serine-dehydrotyrosine-

40

Luminous organisms (genera)	Luciferins, other factors	Luciferases (kDa)	Emissiom max (nm)
Bacteria (Photobacterium; Vibrio)	$FMNH_2$; RCHO	80	495-500
Dinoflagellates (Gonylaulax; Pyrocystis)	Tetrapyrrole; H^+	60-130	475
Cnidarians (Aequorea, Renilla)	Coeletterazine; Ca^{2+} (imidazopyrazine nucleus)	21	460-490
Annelids (Diplocardia)	N-isovaleryl-3-amino propanal, H_2O_2	300	500
Molluscs (Latia)	Enol formate; terpene or aromatic aldehyde	170	500
Crustacea (Vargula; Cypridina)	Imidazopyrazine nucleus	60	465
Insects (Photinus; Photuris)	(Benzo)thiazole; ATP; Mg^{2+}	100	560

Table 1 c   *Chemistry and colour of bioluminescence in different organisms*

glycine derivative that forms the chromophore (Prasher 1995) (figure 1.9). The peptide sequence concerned (serine-tyrosine-glycine) is found in other proteins, but does not undergo the chemical transformation needed to form the pigment. This chemical transformation is an autocatalytic oxidation that can occur so long as most of the GFP polypeptide sequence is expressed and oxygen is present. The green fluorescence requires further interaction of this chromophore with other parts of the protein. No prosthetic group needs to be added. The primary structure requirements for the fluorescent activity of GFP are relatively extensive and this is compatible with the view that the primary structure serves an autocataltic function. Deletion analysis of the GFP protein has shown that amino acids 2-232 of a total 238 amino acids are required for the characteristic emission and absorption spectra of the protein and for chromophore formation (Dopf and Horiagon 1996). The three dimensional structure of GFP has recently been solved (Ormo, Cubitt et al. 1996) and looks remarkably as if it had been engineered for use as a protein tag (figure 1.8). Most of the polypeptide is folded into a tight, cylindrical barrel structure consisting of a highly regular series of 11 β-strands. The outer diameter of the barrel is about 30 Å and the inner surface is quite polar, accommodating a single α-helix. The barrel is capped on both top and bottom, so the chromophore is packaged in an environment entirely composed of protein, completely protected from the surroundings. GFP thus has all the right properties to make it a useful fluorescent tag. Detection of GFP in living cells only requires excitation by light at 395 nm or 470 nm. The protein itself does not appear to be cytotoxic, except when expressed in large quantities. In contrast, the assay of GUS expression is cytotoxic (Jefferson 1989) and firefly luciferase requires the substrate luciferin (Ow, Wood et al. 1986; Millar, Carre et al. 1995). Another advantage of GFP is that it is relatively small (29.9 kDa) and can tolerate both N- and C- terminal fusions, lending itself to studies of protein localisation and intracellular protein trafficking (Kaether and Gerdes 1995).

The first report of the use of transgenic GFP as a reporter for gene expression was in *Caenorhabiditis elegans* (Chalfie, Tu et al. 1994). The nematode was stably transformed with a GFP expression construct in which the GFP coding sequence was placed under the control of the promoter for the *mec-7* β-tubulin gene, expressed in only a subset of neurones. In worms expressing this regulated GFP, the pattern of fluorescence *in vivo* was similar to that previously seen with antibodies against Mec-7 protein. GFP made it possible to see the cell bodies and the process of growing neurones. The GFP-expressing worms went on to become viable adults, indicating that GFP is not toxic. GFP has now been successfully used in a variety of systems including *Drosophila* (Brand 1995), *Dictyostelium* (Hodkinson 1995), yeast (Stearns 1995), mammalian cells (Pines 1995) and more recently, plants (Chiu, Niwa et al. 1996). GFP has been targeted to organelles such as mitochondria (Rizzuto, Brini et al. 1995), chloroplasts (Chiu, Niwa et al. 1996), nuclei (Grebenok,
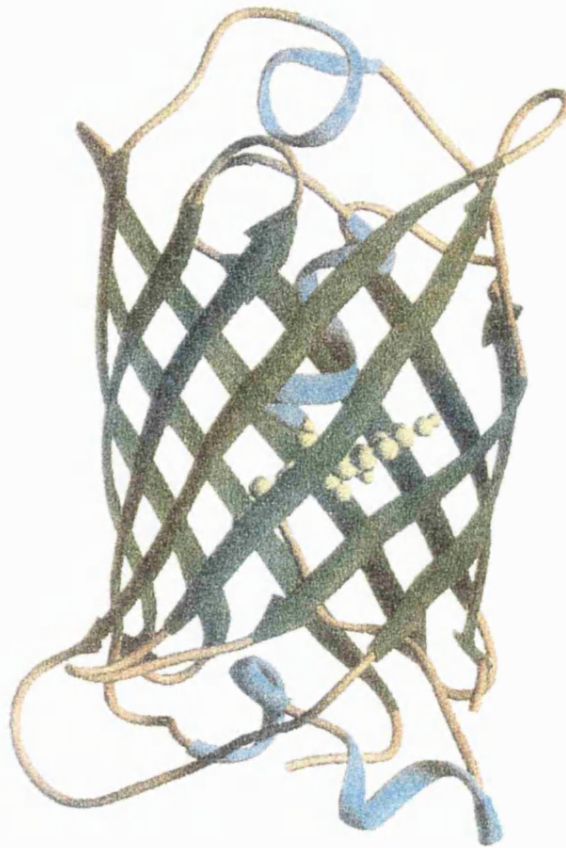
Figure 1.8 *The three dimensional structure of wild-type GFP, adapted from Yang et al. 1996. The 11β-strands comprising the barrel are shown in green; the caps at the top and bottom of the barrel are shown in brown; the chromophore (yellow) is found in the middle of the barrel along the central helix shown in blue.*
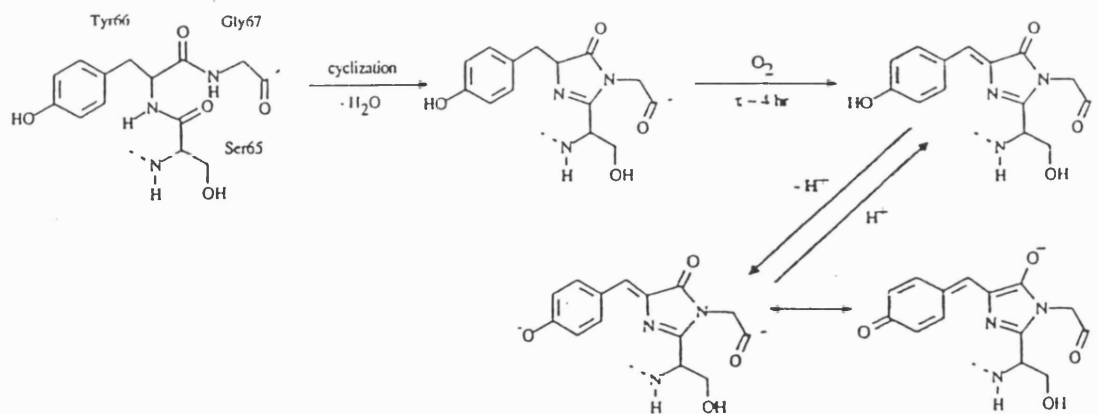
Figure 1.9 *Proposed biosynthetic scheme for the chromophore of GFP. From Heim et al. 1994.*

43

Pierson et al. 1997) and the secretory apparatus (Kaether and Gerdes 1995) where it becomes fluorescent and allows visualisation of organelle dynamics *in vivo*.

1.4.2 Spectral variants of GFP

Despite the many early successes using this reporter, wild-type GFP is still not suitable for many applications. Reasons include low fluorescence intensity and a significant lag in the [poor solubility*] development of fluorescence after protein synthesis. Mutants of GFP have been created with shifted emission and excitation spectra (Delagrave, Hawtin et al. 1995; Ehrig, O'Kane et al. 1995), enhanced emission intensities and rapid chromophore formation by mutation of the amino acids that make up the chromophore (Heim, Cubitt et al. 1995). For example, changing serine 65 to threonine creates a mutant form of GFP whose excitation spectra is shifted to 488 nm and the fluorescence is found to be 35x brighter. The expression of several genes or the co-localisation of several protein fusions can now be simultaneously assayed in a single cell transformed with multiple spectrally distinct forms of GFP as reporters (Yang, Kain et al. 1996b).

1.4.3 Codon optimisation of GFP

In higher eukaryotes, problems with poor expression of GFP have been attributed to unsuitable codon usage or lack of the correct gene regulatory elements. To improve on these qualities, GFP coding sequences have been re-engineered to contain codons preferentially found in the organism of choice. The first example of this is the construction of a unique GFP variant for expression in human cells which contains a chromophore mutation (Serine 65⇒Threonine) and is codon-optimised for higher expression in mammalian cells (Zhang, Gurtu et al. 1996). Optimisation of codon usage was achieved through DNA shuffling or 'Molecular Evolution', a technique for *in vitro* recombination of pools of homologous genes. Coupled with selection or screening, this homologous recombination process is the most effective known process for combining positive mutations and simultaneously removing negative mutations from the sequence pool. The pool of genes is fragmented into random size pieces, and the PCR reassembly of full-length genes from the fragments via self-priming yields cross-overs due to PCR template switching (Stemmer 1994). These changes in the GFP coding sequence provide an enhanced GFP (EGFP) that greatly increases the sensitivity of the reporter protein for human cells.

1.4.4 Expression of GFP in plants

In plants, GFP has been successfully used as a reporter for both moncots and dicots (Haseloff and Amos 1995; Heinlein, Epel et al. 1995; Sheen, Hwang et al. 1995; Chiu, Niwa et al. 1996). The work of Haseloff et al. has shown that lack of expression in

44

---

*Solubility
Another reason for poor fluorescence and possible toxicity of GFP is the formation of GFP aggregates. "Molecular Evolution" through DNA shuffling experiments have resulted in improved GFPs with three mutations which enhance solubility in the cytosol (Crameri, Whitehorn et al. 1996).

Arabidopsis is due to biased codon usage and also to the presence of a cryptic intron in the GFP coding sequence. By removing the splice site and favouring *Arabidopsis* codon bias (low A+U content), a fluorescent GFP can be stably expressed in this plant (Haseloff and Amos 1995). Similarly, enhanced expression of GFP in tobacco has been attained thorough modification of codon usage and the removal of a cryptic intron (Rouwendal, Mendes et al. 1997). In higher plants, splicing may be circumvented using GFP fusions with RNA viruses (e.g. potato virus-x), as viral replication is confined to the cytosol (Oparka, Roberts et al. 1995). Visualisation of the replication and spread of plant viruses in their host has been achieved through fusion of GFP DNA with viral genes (Epel, Padgett et al. 1996). Incorporation of a plant intron into the coding region of GFP has also yielded an improvement in expression levels (Pang, DeBoer et al. 1996). A further feature of GFP as a reporter molecule in plants is that it can be induced to investigate promoter specificity. Tobacco plants were stably transformed with an *Arabidopsis* drought-inducible promoter or a light-inducible promoter fused to GFP and subject to water stress or light stimuli respectively, leading to GFP accumulation (Chiu, Niwa et al. 1996).

A desirable goal for molecular biological techniques in *Chlamydomonas* is the development of a suitable reporter molecule for gene expression in living cells. GFP has been used as a reporter in a variety of heterologous systems, and since its expression appears to be species independant, GFP could prove to be extremely powerful as a reporter of gene expression and protein localisation for the alga.

1.5 Negative selectable marker for homologous transformation

1.5.1 The need for a means of gene targeting in *Chlamydomonas*

Due to the low rate of homologous recombination in *Chlamydomonas*, it has been difficult for researchers to target genes of interest to create site directed or null mutants. It is desirable to find a means of selecting for rare homologous recombination events, and such an approach would greatly increase the value of the organism for research. Similar problems have hampered targeted gene disruption in mammalian cells. In these cells the frequency of non-homologous recombination may be 1000 times that of targeted integration. A very effective 'positive-negative' selection has been developed to identify targeted gene disruptants (Mansour, Thomas et al. 1988; Capecchi 1989).

1.5.2 A strategy for development of a selectable marker for homologous recombination events

In this method, mammalian cells were transfected with a plasmid that contains two selectable marker genes: *neo*, which confers resistance to geneticin (G418), and *HSV-tk* (Herpes Simplex Virus thymidine kinase) which causes sensitivity to gancyclovir (GANC). The disruption construct contains a *neo* gene flanked by DNA sequences that are homologous to

the genomic target site. Homologous recombination integrates the *neo* gene into the genomic target site and generates G418 resistant cells. Random integration of the *neo* gene will also produce G418 resistant cells. These can be eliminated by flanking the homologous DNA sequences with a *HSV-tk* gene as a negative selectable marker. Targeted integration results in the loss of the *HSV-tk* gene, producing cells that survive double selection on G418 and GANC.

Gene targeting via negative selection thus involves the use of a transformation construct in which a lethal gene is inserted downstream of the target sequence - usually a mutated version of an endogenous gene disrupted with a positive marker that confers drug resistance on the cell. The lethal gene will have its own promoter and regulatory elements and if integrated into the genome, will be activated, causing cell death. If homologous recombination occurs, only the target DNA will integrate, the remaining vector will be degraded and the cell will remain viable, able to survive both positive and negative selection. In the case of random integration of the transformation construct, the entire vector will integrate including the target sequence and the lethal gene, resulting in cell death. Please refer to figure 5.1 on page 131.

1.6 The use of antisense as a means of down-regulation of specific genes

The use of antisense technology has now taken a firm hold in attempts to manipulate eukaryotic gene expression. Expression of antisense RNA is achieved by using part of the gene of interest to construct a transgene in which the non-coding DNA strand is transcribed. Antisense transformation can result in a dramatic decrease in expression of the homologous host gene and has provided a powerful tool for establishing the function of plant genes by down regulating a specific gene or gene family of interest (Mol, Blockland et al. 1994). Antisense regulation also has commercially important applications in crop improvement (Gasser and Fraley 1989; Fray and Grierson 1993; Bourque 1995). Additionally, antisense technology has wide implications for clinical medicine, the discussion of which is outside the scope of this thesis (Woolf 1996).

Antisense RNAs were initially recognised in bacteria as naturally occurring mechanisms for the regulation of gene expression (Simons 1988). The first engineered experiments in this area were done in mouse cells expressing complementary RNAs against thymidine kinase (Izant and Weintraub 1984), but plants were the first higher multicellular eukaryotic organisms in which artificial antisense genes were tested. The first report on artificial antisense regulation of gene expression described the transient inhibition of chloroamphenicol acetyl transferase (CAT) activity in carrot cell cultures (Ecker and Davis 1986). A sense *cat* gene was introduced into protoplasts by electroporation, giving rise to detectable CAT enzyme activity. Co-electroporation with an antisense *cat* gene reduced CAT activity by up to 95%. The first wild-type plant gene artificially regulated by an antisense gene was the chalcone synthase endogenous gene (*CHS*) from petunia and tobacco plants,

expressed in flower tissue. Transformed plants showed an altered pigmentation of the flower with high frequency (Krol, Mol et al. 1988). Inactivation of another plant gene encoding polygalacturonase (PG) in tomatoes paved the way for the use of antisense in crop improvement and the first major contribution of antisense technology to the agriculture industry. PG antisense RNA expressed in transgenic tomato was shown to cause a reduction in PG enzyme activity as well as PG mRNA during fruit ripening. Subsequent studies of the control of cell-wall metabolism and texture change during ripening through antisense technology have lead to the application of antisense technique to extend the shelf-life of the fruit (Fray and Grierson 1993).

The exact mechanism of antisense inhibition remains unclear, but the rationale behind the application of antisense genes is the *in vitro* synthesis of complementary RNA, which subsequently hybridises to its target RNA and prevents its translation. Inhibition is restricted to the target gene (and perhaps related genes with high sequence homology). Varying levels of inhibition are seen in different transformants carrying identical transgenes: this leakiness allows for the observation of phenotypic effects of genes whose complete inhibition would be lethal.

As knowledge of this technology broadens, it is apparent that antisense strategies can provide powerful insights into plant molecular genetics at every level. Antisense has been used in the production of deficiency mutants. Antisense RNA can be a means of partially or completely repressing the synthesis of a gene product in attempts to observe steps in a metabolic pathway and to ascertain the relationship between transcript and protein levels. In whole plant systems antisense has been utilised in the determination of sequence or promoter specificity in different tissue types, crop improvement and the unravelling of the genetics of plant development (Bourque 1995). Such an approach would be extremely useful in *Chlamydomonas* as a means for down-regulation of specific genes without the need for gene targeting.

1.7 Aims of this research

This thesis presents the results of investigation into possible new tools for nuclear transformation:

i) The results of a pilot study of analysis of Expressed Sequence Tags from *Chlamydomonas* are presented. Sixteen cDNAs chosen at random from a *Chlamydomonas* cDNA library were sequenced. The sequences were analysed for sequence similarities to other organisms and for functional motifs. A further 84 clones were then subject to EST analysis.

ii) The use of Green Fluorescent Protein as a reporter molecule for *Chlamydomonas* was investigated. The GFP coding sequence was introduced into the nuclear genome of *Chlamydomonas* under control of *Chlamydomonas* gene regulatory elements and the transformants examined for expression of Green Fluorescent Protein. *Gfp* was also introduced into the chloroplast DNA of *Chlamydomonas* in an attempt to gain stable expression of GFP in the chloroplast.

iii) Attempts were made to develop a negative selectable marker for homologous transformation based on the creation of mutants defective in uracil or acetate utilisation and their restoration to wild-type phenotype by transformation with the wild-type gene.

iv) The use of antisense expression in *Chlamydomonas* was investigated. An antisense section of the *Chlamydomonas* nuclear gene encoding oxygen evolving complex 1 (*oee1*) was inserted into the genome of *Chlamydomonas* under the control of a strong *Chlamydomonas* promoter and the resulting transformants analysed for phenotypic effects and the presence of an antisense transcript.

Chapter 2
MATERIALS AND METHODS

Chapter 2: Materials and Methods

2.1 Materials
2.1.1 Reagents and enzymes
Restriction endonucleases and other modifying enzymes and polymerases were purchased from New England Biolabs (Hitchin), Stratagene (Cambridge), Promega (Southhampton) and Boehringer Mannheim (Lewis).

All common laboratory reagents were of the highest grade available, and where no supplier is mentioned, were purchased from Sigma Chemical Co. (Dorset).

2.1.2 Radioactive compounds
Radiolabelled dNTPs: $[\alpha\text{-}^{32}P]dCTP$, $[g\text{-}^{32}P]dCTP$ and $[\alpha\text{-}^{35}S]dATP$, were purchased from Amersham International.

2.1.3 Oligonucleotides
Oligonucleotides were custom synthesised by PE-Applied Biosystems, UK (Cheshire).

2.1.4 Solutions and Media
Details of growth media for bacterial strains are summarised in Table 2 a. Details of growth medium for algal strains are summarised in Table 2 b. Where neccessary, growth medium was solidified with 2% bactoagar (Difco).

Solutions and media were sterilised by autoclaving at 15 psi, 121 °C for 20 minutes, or in the case of heat sensitive reagents, passed through a 0.2μm pore-diameter filter.

2.2 Organisms and their growth
2.2.1 Bacterial strains

The *E.coli* strains DH5α phenotype: $F^-_{,\emptyset}$ 80d*lac*ZΔM15,Δ(*lac*ZYA-*arg*F), U169, *deo*R,

*rec*A1, *end*A1, *pho*A, *hsd*R17(r_{K-} , m_{K+}), *sup*E44, λ^- , *thi*-1, *gyr* A96, *rel* A1 and JM110

phenotype: *rps*L (Str$^+$) *thr leu thi-1 lac*Y *galK gal*T *ara ton*A *tsx dam dcm sup*E44 Δ(*lac-proAB*) [F' *tra*D36 *pro*AB *lac*I^{q}ΔM15] were used for the propagation of plasmids. The *E.coli* strain C600 (*sup*E44, *hsd*R17, *rec*A1, *end*A1, *gyr*A46, *thi rel*A1, *lac*-F'[*pro* AB+ *lac*Iq *lac*ZΔM15 Tn10(tet^r)] was used to propagate λgt10.

All bacterial strains are available from Stratagene.

Table 2a   E. coli *growth medium*

For 1 litre	LB
Bactotryptone	10g
Yeast Extract	5g
NaCl	10g
dH_2O	to 1 litre

Table 2b   Chlamydomonas *growth medium*

For 1 litre	TAP medium	Tris minimal medium
dH_2O	975ml	975ml
Tris	2.42g	2.42g
4x Beijerink salts*	25ml	25ml
1M (K) PO_4 pH 7.0#	1ml	1ml
Trace elements+	1ml	1ml
Glacial Acetic acid	~1ml to pH 7.0	-
Concentrated HCl	-	to pH 7.0

*4x Beijerink Salts

    16g NH_4Cl (omitted in TAP -N medium)
    2g $CaCl_2$
    4g $MgSO_4$
    dissolve in 1 litre distilled water

#1M (K)PO_4

    250ml 1M K_2PO_4
    70 ml 1M KH_2 PO_4
    adjust to pH 6.9 with KOH

+Trace Elements

    i) Dissolve in 550ml distilled water 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 $FeSO_4.6H_2O$
        1.61g $CoCl_2.4H_2O$
        1.1g $(NH_4)6Mn_7O_{24}.4H_2O$

    ii) Dissolve 50g $EDTA.Na_2$ in 250 ml H_2O and add to the above solution. Reheat to 100°C. Cool to 80°C and adjust to pH 6.5 with 20% KOH.

    iii) adjust to 1 litre and allow to stand at room temperature for two weeks to form a rust coloured precipitate. The solution will change from green to purple.

    iv) Filter through three layers of Whatman No. 1 paper under suction until the solution is clear. Store at 4°C.

51

2.2.2 Growth and maintenance of bacterial strains

Stock *E. coli* strains were maintained on LB medium solidified with 2% (w/v) bacto agar at 4 °C. Stock plates were restreaked to fresh plates every 4-6 weeks, grown overnight at 37 °C, then placed in the fridge. Long term storage of *E. coli* cultures was in the form of frozen glycerol stocks at -70 °C.

2.2.3 Algal strains

All algal strains used are maintained in the Purton laboratory and were originally obtained from Dr. E. Harris at the *Chlamydomonas* Genetics Center at Duke University, North Carolina.

Natural wild-type isolates of *C. reinhardtii* used are CC1021 (mt+) and CC2290 (mt-) strain. A wild-type isolate of *Chlamydomonas smithii* , CC1373 (mt-), which is interfertile with wt *C. reinhardtii* was also used.

C. reinhardtii strains used in transformation protocols are the cell-wall deficient, arginine requiring strain *arg7-8* cw_d (mt-), CC-363, and the cell-wall deficient cw_d 10 (mt-), CC-849.

Other strains detailed in this thesis are the oxygen evolving enhancer 1 (OEE1) deficient, cell-wall deficient strain, FuD44 cw_d, and a non photosynthetic mutant ($\Delta psbH$) created in this laboratory from the wt strain CC-1021 via disruption of the *psbH* gene with the *aad*A cassette, conferring spectinomycin resistance (O'Connor, Ruffle et al. 1998).

2.2.4 Growth and maintenance of algal strains

Chlamydomonas strains were maintained on tris-acetate phosphate (TAP) medium (Gorman and Levine 1965) solidified with 2% (w/v) bacto agar. Stock plates were restreaked to fresh plates every 4-6 weeks and maintained under a photon flux of 20 $\mu E/m^2/s$ at 18 °C. Before attempting any manipulation of the strains (e.g. DNA extraction, transformation etc.), cells were restreaked to fresh plates several times over the course of a week and maintained at 45 $\mu E/m^2/s$, 25 °C.

Where an acetate-free growth medium was required, tris minimal medium was used. Tris minimal medium is made in the same way as TAP except that the solution is brought to the correct pH using concentrated hydrochloric acid in place of glacial acetic acid.

Liquid cultures were grown in the appropriate medium in an illuminated orbital shaker (New Brunswick Scientific Co., Model 25) at 80 $\mu E/m^2/s$, 150 rpm and 25 °C. The cell density of *C. reinhardtii* in liquid cultures was accurately determined by aseptically removing a 1 ml sample and adding 10 μl tincture of iodine (0.25 g iodine in 95% ethanol) to immobilise the cells which were then counted using a haemocytometer.

Where cells were required to be grown in low light or complete darkness (e.g. in the case of photosynthetic mutants) this was achieved by covering flasks or plates with 3MM paper or with foil.

2.2.5 Yeast strains

Schizosaccharomyces pombe leu 1-32 h⁻ transformed with Green Fluorescent Protein wild-type DNA was kindly donated by Dr. K. May of the Biology Department, UCL.

2.3 General molecular biology techniques

2.3.1 Restriction endonuclease digestion

DNA was digested with restriction endonucleases using approximately 2 units of enzyme per µg of DNA. Buffer conditions and incubation temperatures were as recommended by the enzyme suppliers.

2.3.2 Dephosphorylation of DNA

Calf intestinal phosphatase was added to DNA digestions at 0.1 unit/pmol of DNA ends. The phosphatase was inactivated by heating to 80 °C for 10 minutes.

2.3.3 Ligation of DNA

For ligation of DNA involving cohesive termini, reaction mixtures (20 µl) included up to 4 µg insert DNA , 0.5 µg of vector DNA and 1 unit of T4 DNA ligase in buffer supplied by the enzyme manufacturer. Ligations were carried out at 18 °C for three hours or overnight. For blunt ended ligations, more DNA was included with up to 5 units of T4 DNA ligase, and the reactions carried out at 12 °C for 3 hours or overnight.

2.3.4 Agarose gel electrophoresis of DNA

DNA fragments were separated by electrophoresis through agarose gels (1% w/v) cast in 0.5xTBE buffer (45mM Tris-HCl, 45mM boric acid, 1.25mM EDTA.Na₂). Ethidium bromide was included in gels at 0.5 µg/ml. Samples were mixed at the appropriate ratio mM EDTA.Na₂, 0.01 % (w/v) bromophenol blue, 0.01 % (w/v) xylene cyanol. Gels were run submerged in TBE at 75 V constant voltage for 1-2 hours in a Hoefer 10 cm cooled minigel apparatus. Fragments were visualised under UV light and sized by comparison with standards.

2.3.5 Recovery of DNA fragments from agarose gels

DNA fragments were recovered from agarose gels using a 'geneclean' kit (Stratech) according to the manufacturers protocol and resuspended in TE buffer or sterile ddH₂O.

2.3.6 Transformation of *E. coli* with plasmid DNA

Transformation of *E. coli* strains was performed as follows: a single colony grown for 12-18 h on a 2 % LBA plate was used to inoculate 10 ml of LB. The culture was grown to stationary phase overnight at 37 °C. 100 μl of this was then used to inoculate 10 ml of LB medium and the culture grown at 37 °C with shaking for a further 2.5 hours. Cells were pelleted by centrifugation and resuspended by gentle shaking in 10 ml ice-cold 100 mM $MgCl_2$ then left on ice for 5 minutes after which time they were pelleted once more and resuspended in 1 ml ice-cold 100 mM $CaCl_2$. The cells were left on ice for a further 30 minutes in order to induce competence.

For each transformation, 0.1 ml of competent cells were aliquoted into a 1.5 ml sterile Eppendorf tube containing 10 μl of the appropriate ligation mix, and the transformations kept on ice for 30 minutes so that DNA could precipitate onto the cell surface. The cells were then 'heat shocked' at 42 °C for 90 seconds and immediately returned to ice. 1 ml LB was added and the cells were incubated with constant shaking at 37 °C for 30 minutes to allow for the expression of plasmid encoded antibiotic resistance markers. 100 μl of the cell suspension was then plated out on LB agar plates containing antibiotic and colour selection as necessary. For pUC based vectors, plates contained 50 μg/ml ampicillin and were pre-treated with a 200 mg/ml stock solution of the *lac* inducer IPTG and 40 μl of a 20 mg/ml in dimethylformamide stock solution of the chromogenic substrate X-gal to allow blue-white selection. Cells transformed with pZerO (Invitrogen), a vector that allows direct selection of colonies with inserts via disruption of a lethal gene, were selected on plates containing Zeocin (Cayla, France) at 50 μg/ml and 1 mM IPTG. Controls were included as part of each transformation experiment as follows: cells only, cells transformed with cut and uncut vector, cells transformed with ligated vector.

2.3.7 Isolation of plasmid DNA

Small scale preparations of plasmid DNA ('minipreps') were isolated from transformed *E. coli* using the alkaline SDS method (Birnboim and Doly 1979). 10 ml of LB containing appropriate antibiotic(s) was inoculated with the bacteria containing the plasmid by picking a colony from a plate. The culture was incubated at 37 °C overnight and the bacteria were pelleted by centrifugation. The cells were resuspended in 100 μl of solution I (50 mM glucose, 25 mM Tris-HCl pH8.0, 50 mM $EDTA.Na_2$), incubated at room temperature for 5 minutes and transferred to an Eppendorf tube. 200 μl of freshly made solution II (1% SDS, 0.2 M NaOH) was added in order to lyse the cells. After a 5 minute incubation at room temperature, 150 μl of solution III (2 M acetic acid, 3 M potassium acetate) was added to precipitate cell debris and the samples left on ice for 30 minutes before centrifugation at 13000g. The resulting supernatant was then extracted with an equal volume of phenol:chloroform, following which 0.1 volumes of 3 M sodium acetate pH 6 and 2

54

volumes of ice-cold ethanol were added to the supernatant, and the tube placed at -20 °C for 20 minutes to precipitate nucleic acids. Nucleic acid was pelleted by centrifugation at 13000g for 15 minutes. The DNA pellet was washed with 70 % (v/v) ethanol, then dried under vacuum and resuspended in TE/RNaseA.

Larger scale isolation of plasmid DNA was performed using a Qiagen (Dorking) maxiprep kit according to manufacturer's instructions; finally resuspending the DNA in 200 μl of TE pH 8.0. The concentration of the DNA was determined by diluting 2 μl of the concentrated stock into 1 ml of ddH₂O and measuring the absorbance at 260 nm. DNA at a concentration of 1 mg/ml gives a reading of 20 absorbance units at 260 nm therefore the concentration could be calculated and the sample diluted to a working concentration of 1 mg/ml.

2.4 RNA techniques

As RNA is extremely susceptible to degradation and RNases are particularly stable, the following precautions were always taken during RNA manipulation and isolation to eliminate possible RNase contamination. Microbiological, aseptic technique was used at all times to avoid contamination with dust particles, and gloves were worn to prevent RNase contamination from the surface of the skin. All tubes were kept on ice to reduce degradation of RNA by endogenous or residual RNases. Wherever possible, sterile, disposable plasticware was used, and non-disposable plasticware was rinsed with 0.1 N NaOH, 1 mM EDTA followed by RNase free water. Glassware and spatulas were cleaned with detergent, thoroughly rinsed and baked at >240 °C overnight. ddH₂O for resuspension of RNA and making up solutions was treated with 0.1% DEPC and incubated at 37 °C overnight before being autoclaved for 15 minutes to remove DEPC. All dry chemicals were taken from separate stocks which are carefully handled to try to minimise RNase contamination. Finally electrophoresis tanks were cleaned with detergent solution, rinsed with water and dried with ethanol.

2.4.1 Preparation of RNA from *Chlamydomonas*

Isolation of total RNA from *Chlamydomonas* was carried out using RNeasy Spin Columns from Qiagen, essentially via the manufacturer's protocol. Typically, RNA was extracted from 10 ml of cells, pelleted by centrifugation and disrupted via freeze-thawing.

2.4.2 Gel electrophoresis of RNA

A denaturing 1 % agarose gel was prepared by melting 1.2 g agarose in 75 ml DEPC-treated water and 10 ml 10xMOPS buffer (0.2 M MOPS, 80 mM NaOAc, 15 mM EDTA, pH 7.0). The mix was cooled to 50 °C and 16.25 ml of formamide added. Ethidium bromide was also included in the gel at 0.5 μg/ml before casting the gel in a 100 ml casting tray. Samples were mixed at the appropriate ratio with 2x loading buffer (for 1 ml mix 12 μl 10 mg/ml

EtBr, 300 μl 10xMOPS buffer, 80 μl formaldehyde and 300 μl formamide. Store at -20°C). Gels were run submerged in 1xMOPS at 75 V constant voltage for 5-6 hours in a Hybaid midigel apparatus. RNA was visualised under UV light and sized by comparison with RNA standards.

2.5 Filter hybridisation of nucleic acids

2.5.1 Preparation of Southern blots

DNA fragments were separated by agarose gel electrophoresis, stained with ethidium bromide and photographed. The fragments were then transferred to Hybond-N nylon membranes (Amersham International) by capillary blotting.

The gel was immersed in 1.5 M NaCl, 0.5 M HCl for 30 minutes to denature the DNA, rinsed in distilled water and then neutralised for 2 x 15 minutes in 1 M Tris-HCl pH 8.0, 1.5 M NaCl. The gel was again rinsed in distilled water and placed onto three strips of Whatman 3MM paper soaked in 20XSSC (3 M NaCl, 0.3 M sodium citrate) supported on a glass plate. The ends of this paper wick were placed in a reservoir of 20XSSC and strips of Parafilm (Amersham International) placed around all four sides of the gel to prevent the flow of buffer around the gel causing a short circuit. A piece of Hybond-N was placed on top of the gel, followed by a stack of paper towels 5 cm high. Finally the whole assembly was weighted down with a glass plate and a 0.5 kg weight and left overnight at room temperature to allow transfer of the DNA to the nylon membrane by capillary action. The membrane was baked at 80 °C for 1-2 hours in order to immobilise the DNA on the filter.

2.5.2 Preparation of northern blots

Transfer of RNA to nitro-cellulose membrane was carried out essentially as described for DNA transfer, except that the denaturing and neutralising steps were omited. After transfer the membrane was immediately baked for 1 hour at 80 °C to fix the RNA to the nylon support.

2.5.3 Radiolabelling of DNA probes

DNA fragments to be used as probes were labelled using the Prime-It II Random Primer Labelling Kit (Stratagene) using random 10-mer primers and (exo^-) Klenow polymerase, essentially as per the manufacturer's instructions. Once labelled the probe was stored at -20 °C and denatured by boiling for 5 minutes immediately prior to use.

2.5.4 End-labelling of oligonucleotide probes

For probing of plaque lifts, oligonucleotides were end labelled using a method modified from (Sambrook, Fritsch et al. 1989). Reaction mixtures contained 100 ng oligonucleotide, 10 pM [g-^{32}P]dCTP, 20 units of polynucleotide kinase in a volume of 100 μl in 50 mM

Tris-HCl pH 7.5, 100 mM MgCl$_2$, 5 mM dithiothreitol, 0.1 mM spermidine and 0.1 mM EDTA. The end labelling was performed at 37 °C for 45 minutes and the reaction terminated by incubation at 70 °C for 5 minutes.

2.5.5 Hybridisation of DNA probes to membrane-bound nucleic acids

Membranes were blocked in 40 ml of hybridising solution (5xSSPE, 1x Denhardt's solution, 0.5 % SDS and 0.1 mg/ml sheared, denatured salmon sperm DNA) at 65 °C for at least 4 hours with constant turning using Hybaid Bottles and a Hybaid dual hybridisation oven. The pre-hybridising solution was then discarded and a further 20 ml of solution containing the denatured radiolabelled probe added. Hybridisation was carried out overnight at 65 °C with constant turning. Filters were then washed 3x for 20 minutes at room temperature in 2xSSC, 0.1x SDS, then twice at 65 °C for 20 minutes in 0.1xSSC, 0.1 % SDS. After washing, filters were sealed in clingfilm, and autoradiographed at -80 °C for 1-6 days using X-ray film from Kodak (Herts) and intensifying cassettes from Genetic Research Instruments (Dunmow, UK).

2.6 Methods employing recombinant phage

2.6.1 Transfection and plating out of *E. coli* with bacteriophage λ

Cells from an overnight culture of *E. coli* C600 were diluted 100x in LB medium supplemented with filter sterilised maltose to 0.2 %, MgSO$_4$ to 10 mM and grown in an orbital shaker for three hours at 37 °C, 150 rpm. 200 μl of this suspension was then added to 200 μl of 'phage particles at 10^6/ml and left for 20 minutes at 37 °C to allow absorption of the phage. The infected bacteria and these were plated to LB agarose plates (LB + 1.6 % agarose in 13 cm petri dishes) in 6.5 ml of top agarose (LB, 0.2 % maltose, 10 mM MgSO$_4$, 0.7 % agarose) at 42 °C, and incubated overnight at 37 °C. 200 individual plaques were chosen at random and stored in 200 μl SM (10 mM MgSO$_4$, 10 mM tris-HCl pH 7.5) with a few drops of chloroform to prevent further bacterial growth prior to preparation of hi-titre phage stocks.

Hi-titre stocks were prepared for individual cDNA clones by transfection of C600 with individual plaque suspensions as above. The following day, 1 ml of SM was added to the plate surface and left to absorb 'phage particles by gently agitating on a shaking platform for 1-2 hours. The diluent was recovered with a pipette and centrifuged at 4000 g to pellet any bacterial cells or debris. A few drops of chloroform were then added to the supernatant, which was stored at 4 °C.

2.6.2 Preparation of λ DNA by a plate lysate method

DNA was prepared for individual cDNA clones by transfection of C600 with individual plaques as above. The following day, 5 ml of SM λ diluent was added to the plate surface and left to absorb 'phage particles. The diluent was recovered with a pipette and centrifuged at 4000 g to pellet any bacterial cells or debris. The supernatant was transferred to a Corex tube and an equal volume of 20 % polyethylene glycol (PEG) 8000, 2 M NaCl in SM was added. The tubes were left on ice for 1 hour to precipitate the 'phage, which were then pelleted by centrifugation at 10000g for 10 minutes. The 'phage pellet was carefully drained, resuspended in 750 μl LB and transferred to an Eppendorf with 750 μl of a suspension of DEAE cellulose (Whatman) DE52 in LB. The suspension was gently mixed by inverting 20-30 times, followed by centrifugation for 5 minutes. Proteinase K at a concentration on 0.1 mg/ml and SDS to 0.5 % was added to the supernatant in a fresh tube and the mix incubated at room temperature for 5 minutes. 173 μl of 3 M potassium acetate was added (precipitate formed) and the tubes incubated for 20 minutes at 88 °C in order to denature the proteinase (precipitate dissolved) then cooled on ice for 10 minutes (precipitate reformed). Tubes were centrifuged for 10 minutes in a microfuge an equal volume of isopropanol added to the supernatant in order to precipitate the 'phage DNA. The tubes were incubated at -70 °C for 20 minutes, centrifuged to pellet the DNA, which was then washed with 70 % ethanol, dried under vacuum and resuspended in 40 μl TE pH 8.0, ready for restriction analysis.

2.6.3 Plaque lifts and hybridisation with nucleic acid probes

Dry nitrocellulose discs (Amersham) were numbered in pencil and carefully placed onto the surface of the plates. The filters were marked with at least six assymmetrical points using a wide-bore sterile needle, and were lifted off after 30-60 seconds. The filters were placed plaque side up on a tray containing 1.5 M NaCl and 0.5 M NaOH for 1-3 minutes. The filters were then transferred into 1.5 M NaCl, 0.5 M Tris-HCl pH 8.0 for 5 minutes before being rinsed in 2xSSC and allowed to dry at room temperature. They were then baked for 1 hour at 80 °C to bind the DNA to the filter. Prehybridisation and hybridisation conditions were essentially as described in Section 2.5.4. DNA probes were end-labelled as described in Section 2.5.6.

2.7 General *Chlamydomonas* techniques
2.7.1 Small-scale preparation of total DNA from *Chlamydomonas*

Preparation of total DNA from *Chlamydomonas* was by a rapid 'miniprep' method adapted from (Rochaix, Mayfield et al. 1988). A 20 ml culture of *Chlamydomonas* grown to mid-log phase was harvested by centrifugation at 4000 rpm for 5 minutes. The resulting cell

pellet was washed in 1 ml of TAP medium, transferred to a 1.5 ml eppendorf tube and repelleted by centrifugation at 13 000g. The cells were then resuspended in 0.35 ml of TEN buffer (50 mM EDTA.Na$_2$ pH 8.0, 20 mM Tris-HCl pH 8.0, 0.1 M NaCl) with 50 µl of pronase at 10 mg/ml and 25 µl of 20 % SDS and incubated at 55 °C for 2 hours to break down the proteinaceous cell walls. 2 µl of diethylpyrocarbonate (DEPC) was added and incubation continued for a further 15 minutes in the fume hood. The tube was briefly cooled on ice and 50 µl of 5 M potassium acetate added to precipitate debris. The tube was left on ice for a further 30 minutes followed by centrifugation at 13000g for 15 minutes. The resulting supernatant was extracted twice with an equal volume of phenol:chloroform. DNA was subsequently precipitated by adding 0.1 volumes of 3 M sodium acetate pH 6.2, 2 volumes of ice-cold ethanol and left at -20 °C for 20 minutes before centrifuging at 13000g for 15 minutes. The supernatant was removed by aspiration and the pellet washed with 70% (v/v) ethanol prior to being dried in a Rotovac vacuum drier. The DNA was resuspended in 50 µl TE, 0.1 µg/ml RNaseA and left overnight to dissolve. 5 µl of this DNA solution was used for Southern analysis.

2.7.2 Fluorescence screening of *Chlamydomonas*

In order to obtain samples of uniform cell concentration for fluorescence screening in liquid culture, a chlorophyll assay was carried out for each sample. 1 ml of a culture was pelleted by centrifugation and resuspended in 80% acetone. The mixture was then centrifuged once more (12 000 rpm, 1 minute) to remove any cell debris, and the absorbency of the sample at 652 nm measured. The concentration of chlorophyll in the sample is then:

$$\Delta \lambda_{652} \times 2.9 = \mu g \text{ of chlorophyll per ml}$$

The cell concentrations were adjusted accordingly. 3 ml of each cell sample was monitored for fluorescence in a Perkin-Elmer spectrofluorimeter, at the appropriate excitaion wavelength.

Transformants on solid media were assayed for fluorescence using a Fluorescence Video Imaging System. The detection system comprises a TM765 monochrome CCD camera with zoom lens (Pulnix). Data transfer was using SNP-8 frame grabber (Datacell) and Image Analysis was carried out using Optimas 5.0 software (Optimas Corporation). Excitation and emission data was gathered using a Intralux 6000-1 cold-light source and Fibre Optic Ringlight "SMD" (Volpi, Zürich) light source with appropriate filters. For GFP excitation a 480 nm short-pass edge filter was used and for emission a 510 nm narrow band pass filter to filter chlorophyll autofluorescence (Ealing Electro Optics). The antisense transformants

were assayed for fluorescence using 620 nm short-pass edge filter for excitation and a Schott RG665 glass filter.

2.8 Transformation of *Chlamydomonas*
2.8.1 Glass bead method for nuclear transformation

Chlamydomonas cells were transformed by a method based on that of Kindle (1990). Cells to be transformed were grown to early log-phase (1×10^6 cells/ml) in 400 ml of TAP medium in a 1 litre Erlenmeyer flask (see section 2.2.4 for growth conditions). Cultures were then transferred to sterile 200 ml centrifuge bottles and centrifuged at 4000 g for 5 minutes. The resulting cell pellets were resuspended in 1 ml of TAP to obtain a concentrated cell suspension which provides enough cells for approximately 8 separate transformations.

The cell suspension was transferred in 0.3 ml aliquots to sterile 5 ml tubes containing 1-5 μg of the transforming DNA and 0.3 g of 0.4 mm diameter sterile glass beads. Plasmid DNA used to transform *Chlamydomonas* cultures was first linearised using an appropriate enzyme in order to aid integration of the DNA into the genome. Each glass bead/cell/DNA suspension was vortexed for 15 seconds at top speed before addition of 3 ml of molten soft agar (TAP/0.6 % agar) and plating onto selective medium. For direct selection of transformants using the *ble* (phleomycin resistance) and the lac-α marker a slight modification was included in the protocol such that after vortexing the cells were resuspended in 20 ml TAP medium and allowed to grow overnight in the illuminated orbital shaker before being plated to selective medium. This allows the cells to recover and to express the marker gene, *ble*. Transformants were scored and picked to fresh medium after 7-9 days. Integration of DNA into the genome is stable, and once transformed the cells will maintain their new phenotype on non-selective medium.

2.8.2 Biolistic method for chloroplast transformation

Chloroplast transformation of *Chlamydomonas* was performed according to the method of (Boynton, Gillham et al. 1988). Cells were prepared by growing a 25 ml culture of the appropriate culture to stationary phase in dim light, harvesting by centrifugation and resuspending in 0.6 ml TAP and 1 ml of molten top agar (TAP/0.2 % agar). 0.7 ml of this mix was immediately plated onto a TAP plate and spread by swirling. The plates were then stored under dim light. Tungsten particles were prepared by vortexing 60 mg of tungsten powder in 100 μl ethanol and leaving to soak for 15 minutes. The tungsten was pelleted by centrifugation, washed 3x in sterile ddH$_2$O and resuspended in 50 % glycerol such that the final volume was 1 ml. A 25 μl aliquot of the tungsten suspension was transferred to a fresh tube and 2.5 μl transforming DNA (at 1 mg/ml), 25 μl 2.5 M CaCl$_2$ and 10 μl spermidine added with 10 seconds vortexing after the addition of each item. Once the DNA had been

added to the tungsten particles it will start to degrade, so subsequent steps were carried out as quickly as possible. The DNA-tungsten mix was vortexed for 60 seconds, microfuged to pellet the particles and 50 μl of the supernatant removed. The particles were resuspended in the remaining supernatant, and 3 μl of this suspension was loaded into the biolistic device, and delivered to the cells. One tungsten preparation is sufficient to shoot three plates. After shooting, the plates were then left at 18 °C in dim light overnight. 1 ml of TAP was added to the plates, and the cells released by gently stroking with a flamed glass spreader. The TAP was drawn off, mixed with 1 ml molten top agar (TAP/0.7 % agar) and plated onto selective media. Plates were grown in dim light at 18 °C and scored after 5-6 weeks.

2.8.3 Segregation of the chloroplast genome to obtain homoplasmicity

Once transformed colonies were large enough, a loop of cells from the colony was picked, resuspended in 3 ml Tris min and 100 μl of this spread onto a fresh Tris-min plate and grown under high light at 18 °C until single colonies appeared. This procedure was then repeated 5 or 6 times to ensure the transformed cells were homoplasmic.

2.9 Mutagenesis of *Chlamydomonas*
2.9.1 Direct selection

Spontaneous resistance mutants of *Chlamydomonas* were isolated by directly plating cells onto selective medium. Firstly, the level of tolerance of wild-type *Chlamydomonas* to the selective drug was determined by spotting 5 μl of a mid-log phase culture onto TAP plates containing a range of drug concentrations and scoring for growth after 6 days.

A 400 ml culture of wild-type cells was grown to mid-log phase ($\approx 3 \times 10^6$) and the cells counted using a haemocytometer. A 10 μl aliquot of cells was removed and plated directly onto TAP medium to determine the viable cell count, and the remaining cells pelleted by centrifugation, resupended in an appropriate volume of soft-agar, and plated onto a large plate containing the mimimum killing concentration of the drug.

2.9.2 Ultra-Violet radiation

Where resistance mutants could not be recovered by isolation of spontaneous mutants, cells were treated with Ultra Violet radiation, known to induce nuclear mutations (Harris 1989). A 'kill curve' was prepared to determine the irradiation time required to produce a culture where viability is reduced to 10 %. A 15 ml culture of mid-log phase *Chlamydomonas* was placed in a petri-dish with constant stirring using a magnetic stirrer and irradiated from above with a UV light source. 50 μl aliquots of cells were removed at 30 second intervals over the course of 10 minutes. After irradiation, cells were allowed to recover in the dark for 1 hour to prevent photo-reactivation, and then plated directly onto TAP. After 5-6 days, the plates were scored and a graph of viable cells vs. time of exposure to ultra violet radiation

prepared. From this graph, the time of exposure required to produce a 10 % viable culture was determined. A second 15 ml culture of mid-log phase cells was prepared and irradiated to produce a 10 % viable culture. The culture was allowed to recover in the dark for 1 hour, harvested by centrifugation and plated directly onto selective media in soft agar. Colonies were picked to fresh agar after 7-10 days.

2.9.3 Insertional mutagenesis

Mutagenesis of *Chlamydomonas* was also carried out using insertional mutagenesis, utilising the glass bead method of Kindle (see section 2.8.1). The transforming DNA disrupts genes at random by integration into the genome. Cells were transformed using the *lac-α* cassette. Cells to be mutagenised were grown to early log-phase (1×10^6 cells/ml) in 400 ml of TAP medium in a 1 litre Erlenmeyer flask (see section 2.2.4 for growth conditions). Cultures were then transferred to sterile 200 ml centrifuge bottles and centrifuged at 4000 rpm for 5 minutes. The resulting cell pellets were resuspended in 1 ml of TAP to obtain a concentrated cell suspension which provides enough cells for approximately 8 separate transformations. The cell suspension was transferred in 0.3 ml aliquots to sterile 5 ml tubes containing 1 µg of the lac-α DNA and 0.3 g of 0.4 mm diameter sterile glass beads. The tubes were vortexed at full speed for 15 seconds then plated onto selective media. Mutants were transferred to fresh media after 6-10 days.

2.10 Protein analysis
2.10.1 Preparation of protein from *S. pombe*

Proteins were extracted from yeast using a method modified from Booher, Alta et al. (1989). 50 ml yeast cultures were grown to a cell density of 1×10^7 cells/ml at 29 °C. The cells were harvested by centrifugation at 2,500 rpm, 4 °C for 3 minutes and the pellet washed once in 10 ml PBS (phosphate buffered saline; 137 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1mM KH_2PO_4, pH 7.4) then once in 1 ml of PBS before resuspending in 250 µl freshly prepared extraction buffer (25 mM Tris, 80 mM β-glycerophosphate, 15 mM p-nitrophenylphosphate, 20 mM EDTA, 15 mM $MgCl_2$, 1 mM DTT, 0.1 mM NaF, 0.1 % Nonidet P-40, 1mM PMSF (phenylmethylsulphomyl fluoride) and 10 µg/ml each of leupeptin, aprotinin and pepstatin). To break open the cells an equal volume of ice-cold 500 µm acid washed glass beads (Sigma) were added to the cell pellet. This mixture was then vortexed vigorously for 30 seconds between 30 second intervals on ice for 6 minutes. The resulting slurry was centrifuged at 13 000 rpm, 4 °C for 2 minutes and the supernatant

collected. The protein concentration was determined using the Bradford Protein Assay (Biorad).

2.10.2 Preparation of whole cell protein from *Chlamydomonas*

Whole cell extract of *Chlamydomonas* protein was prepared by growing 10 ml cultures to mid-log phase and harvesting by centrifugation. The cells were lysed by freeze-thawing in liquid nitrogen and the cell extract was resuspended in protein loading buffer (312.5 mM Tris-HCl pH 6.8, 50 % (w/v) glycerol, 10 % (w/v) SDS, 0.025 % (w/v) bromophenol blue, 25 % (v/v) 2-mercaptoethanol). Insoluble material was removed by centrifugation and the mixture boiled for 5 minutes prior to loading the gel. The protein concentration was determined using the Bradford Protein Assay (Biorad).

2.10.3 Tris-Tricine SDS-PAGE

Tris-Tricine SDS-PAGE was performed according to the method of Schägger and von Jagow (1987) using a Biorad Mini Protean II protein gel apparatus. The protein resolving gel was made by mixing 17.5 % polyacrylamide (acrylamide:bis-acrylamide ratio of 29:1), 1 M Tris-HCl pH 8.45, 0.1 % SDS and 13.3 % (w/v) glycerol. The mix was polymerised chemically by adding 0.025 % TEMED and 0.025 % AMPS. The stacking gel was made up of 4.2 % polyacrylamide, 7.4 M Tris-HCl pH 8.45 and 0.07 % SDS, again polymerised by the addition of AMPS and TEMED. The anode buffer was 0.2 M Tris-HCl pH 8.9 and the cathode buffer was 0.1 M Tris pH 8.25, 0.1 M tricine, 0.1 % SDS. Samples were electrophoresed at 10 milliamps overnight. Bio-rad medium-range prestained markers were used as molecular weight standards.

2.10.4 Transfer of protein to nitrocellulose membrane

Proteins separated by SDS-PAGE were blotted onto Hybond nitrocellulose membrane using a semi-dry trans blotter (Biorad) by the method of Towbin, Staehelin et al. (1979). The transfer buffer was 20 % (v/v) methanol, 25 mMTris, 192 mM glycine, 0.1 % SDS, and transfer was performed at 25 V for 1 hour.

2.10.5 Immuno-detection of protein

Filters were blocked with 5 % low fat skimmed milk powder in TTBS and probed with the appropriate antibody at the correct dilution. For detetion, ECL Western blotting reagents were used (Amersham International) essentially as described in the manufacturer's protocol.

2.11 Miscellaneous techniques
2.11.1 The polymerase chain reaction

Amplification of DNA fragments by PCR was performed in a Techne Progene Thermal Cycler using *Vent* DNA polymerase from Promega. Reactions contained 10 ng plasmid DNA or 1 μg genomic DNA , 50 μM each dATP, dCTP, dGTP and dTTP, I unit *Vent* DNA polymerase, 1 μM each oligonucleotide, 1x Vent reaction buffer (Promega) and made up to volume with sterile ddH$_2$O. Conditions of the reaction cycles were determined by the length and composition of the oligonucleotide primers and by the predicted length of the product. Individual PCR primers are detailed in the relevant sections of each results chapter.

2.11.2 DNA sequencing

DNA sequence analysis was performed by a procedure derived from the di-deoxy chain termination method of Sanger, Nicklen et al. (1977) using a Sequenase II sequencing kit.

10 μg of plasmid DNA, made up to a volume of 18 μl with ddH$_2$O was used as template and prepared by adding 2 μl of 2 M NaOH and heating to 68 °C for 20 minutes to denature the DNA strands. The DNA was precipitated by adding 10 μl of 3 M sodium acetate pH 6.0, 100 μl of absolute ethanol and incubating at -20 °C for 30 minutes. The DNA was pelleted in a microfuge, washed with 70 % ethanol, dried under vacuum and resuspended in 7 μl ddH$_2$O. For the seqencing reactions, 1μl of oligonucleotide primer at 40 ng/μl and 2 μl of sequenase reaction buffer were added to the denatured primer. The mix was heated to >65 °C for 2 minutes and allowed to cool slowly to <35 °C to allow annealing of the primer. Sequencing was then performed according to the manufactuer's protocol. The sequencing reactions were heated to 80 °C for 2 minutes and half of each sample loaded onto a 6 % polyacrylamide (acrylamide:bis-acrylamide, 19:1), 7 M urea, 1x TBE gel (Scotlab-Strathclyde) which had been polymerised chemically using 0.025 % TEMED and 0.025 % ammonium persulphate. The gel was run under 35 V constant voltage using a 30 cm vertical gel apparatus (Cambridge Electrophoresis). Gels were fixed in 10 % methanol, 10 % acetic acid and dried onto Whatman 3 MM paper (Bio-Rad Gel Drier, 80 °C), and autoradiographed for 1-6 days.

Additional automated dye-termination sequencing was carried out by Ms Laura Winskill using a Perkin-Elmer ABI Genetic Analyser 310.

Chapter 3

GENERATION OF EXPRESSED SEQUENCE TAGS (ESTS) FROM *CHLAMYDOMONAS REINHARDTII*

Chapter 3:

Generation of expressed sequence tags (ESTs) from *Chlamydomonas reinhardtii*

3.1 Introduction

An Expressed Sequence Tag (EST) is part of a sequence from a cDNA clone that corresponds to an mRNA. Automated single read sequencing from one or both ends of cDNA clones to generate ESTs is an efficient strategy for identifying new genes and characterising the expressed gene content of cells, tissues and species. ESTs are generated by partial sequencing of randomly chosen clones from a cDNA library and subsequently analysed by comparison of the derived protein sequences with existing sequences in the protein databases. Comparison of this data with other logged sequences yields matches to known sequences where the gene product has been previously identified, or to other ESTs. The translated data may be examined for functional motifs such as zinc fingers, giving a clue to the function of the gene product, or aid in identification by similarities to known proteins. ESTs can be used in the construction of complex physical maps, and this technique is now being exploited widely in large-scale sequencing projects such as the Human Genome Project (Wolfsberg and Landsman 1997). A complete EST database and clone collection for a model organism for which well developed genetic, transformation and antisense inactivation methods are available will aid the functional identification and characterisation of new genes isolated by sequencing methods. Physical mapping of ESTs by hybridisation to cosmids or YACs can yield candidate phenotypes and genomic clones suitable for use in transformation rescue of mutant phenotypes (McCombie 1992). ESTs can be mapped back to their origin in the genome either by hybridisation or by PCR based approaches. If suitable primers to an EST are designed, PCR can be used to amplify the corresponding sequence from genomic DNA, converting the EST to a Sequence Tagged Site which can be mapped to a location in the genome. ESTs are therefore useful both as a source of novel genes and as markers for genetic maps.

In order to obtain a full inventory of all the genes expressed in a whole organism it is necessary to use cDNAs from different tissue types. In the human genome, for example, the gene expression profile for the liver will differ markedly from that of the brain, so in order to capture all cDNAs expressed it is necessary to use cDNA libraries prepared from different tissue types. An extension of this process comes with Differential Expression Analysis - it is possible to identify gene products expressed only in the brain, for example, by comparison with cDNA profiles of other tissue types. Similarly, for plant whole genomes, cDNA libraries may be selected from different tissue types e.g. in *Arabidopsis* an EST project encompassing five cDNA libraries prepared from five different tissue types was undertaken.

In order to identify cDNAs corresponding to mRNAs expressed in dividing cells, libraries were constructed from suspension cultured cells as well as from leaf strips cultured *in vitro* to induce cell division. Libraries from flower buds and developing siliques were used to sequence cDNAs prepared from mRNAs corresponding to flower development and embryogenesis respectively. For the study of genes expressed during hypocotyl elongation, libraries were constructed from 5-day old etiolated seedlings (Hofte 1993).

The development of cDNA microarrays prepared by high speed robotic printing of cDNAs onto glass for subsequent analysis has allowed the automation of this process (Schena, Shalon et al. 1995). The technology is based upon a combination of light-directed synthetic techniques (photolithography) developed in the semiconductor industry with conventional oligonucleotide synthetic chemistry allowing the automated production of up to 100 000 specific cDNA sequences on the surface of a microchip in only a few hours. Such chips are considered one of the most promising of genome technologies. These 1.0 cm^2 DNA chips can be used to quantitatively monitor differential expression of genes using a highly sensitive two-colour hybridisation assay. Array elements that display differential expression patterns can then be characterised by sequencing (Schena 1996). *Chlamydomonas* has only one tissue type. However, comprehensive libraries of mutants, growth of the alga under different limiting conditions and the ability to synchronise cell cultures mean that a differential expression analysis project would be useful as a means of detecting genes expressed under varying growth conditions, looking at stress related proteins for example and in the identification of genes expressed at different stages of development, e.g. during gametogenesis.

The entire genome sequence of several organisms have been elucidated by sequencing efforts and this data enhances ongoing research efforts and opens new areas of research. In order for *Chlamydomonas* to remain an attractive model system, the identification and sequencing of all of its genes is a desirable goal. A prerequisite for this study is to develop a robust and routine method for the generation and analysis of ESTs in *Chlamydomonas*. This chapter reports the results of a pilot study on the generation of ESTs for *Chlamydomonas*.

3.2 Results

3.2.1 Generation of individual clones from a cDNA library

A cDNA library constructed in the bacteriophage λgt10 was kindly donated by Lars Gunner-Franzèn (University of Stockholm). The library was constructed using asynchronous vegetative cells of strain cw15 (mt⁻). The cells were grown in aqueous TAP medium in continuous light and harvested at low density (approximately 10^6 cells per ml). The library has 2x10^6 primary clones and had been amplified once.

E. coli C600 was transfected with the library, and 200 individual plaques picked, corresponding to 200 individual cDNAs. High-titre λ stocks were prepared for 100 of the clones with the intention of excising the insert DNA from its λgt10 arms and cloning the cDNA into a vector suitable for sequencing. This approach proved to be time consuming, so in order minimise time spent on preparation of DNA templates for sequencing a simple and rapid PCR procedure was tested as a means of generating templates for sequencing using primers raised to the λgt10 arms.

3.2.2 Generation of ESTs by PCR from high-titre phage stocks

The PCR reaction was inherently problematic due to the G+C rich nature of the *Chlamydomonas* nuclear DNA. G+C rich DNA has a high melting temperature and so high temperatures are required to dissociate the duplex. Primers with a high T_m were raised to areas of the λ arms with a high G+C content. Denaturing and annealing of the primers could then be carried out at high temperature so that the template remained disassociated prior to the binding of the polymerase enzyme (figure 3.1). A highly heat stable DNA polymerase, *Vent* DNA polymerase (Promega) was used to this end. *Vent* also has proof-reading ability for accurate reproduction of sequence. The PCR cycling times and temperatures are shown in figure 3.2.

Initially PCR products could not be obtained by high temperature PCR using the λgt10 high temperature primers. Promega recommend that the concentration of Mg^{2+} ions in a PCR reaction is critical when using α DNA polymerase as Mg^{2+} ions are required as a cofactor for the enzyme. The concentration of Mg^{2+} was titrated between 1 μM and 6 μM to find a suitable working concentration. A 2 μM [Mg^{2+}] was found to produce a PCR product (figure 3.3). The amplification products were analysed by agarose gel electrophoresis (figure 3.4) to exclude those cDNAs that had very short inserts, lacked inserts completely or failed to produce a PCR product. In this way 29 of the 100 candidate cDNAs that had been made from high-titre phage stocks could be eliminated, optimising the sequencing effort.

3.2.3 Generation of ESTs by PCR directly from phage plaques

As the preparation of high-titre phage stocks is also a time consuming process, it is desirable to be able to produce cDNAs from a plaque picked directly from the plated out library. A PCR protocol was developed that would allow the phage DNA to be amplified directly from low-titre phage stocks. A plaque was picked and resuspended in 100 μl of 10

λgt10 arms   Insert DNA   λgt10 arms

cag agc tgc cat tga gca gcc agt c   ggc ttt ggt gcc agt tcg ggt ttg c
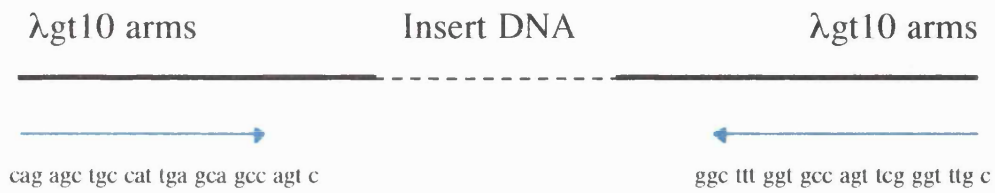
- Insert DNA is GC rich ∴ duplex has a high melting temperature

- Primers designed to λgt10 arms with high GC content ∴ high annealing temperature

$$T_m \text{ Primer } 1 = 80°C$$

$$T_m \text{ Primer } 1 = 82°C$$

- PCR reaction carried out at high temperature i.e.

Denaturing performed at 98°C
Annealing and extension performed at 72°C
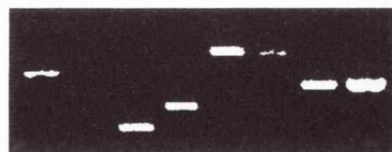
Random PCR products for sequencing

Figure 3.1 *PCR strategy for the generation of ESTs*
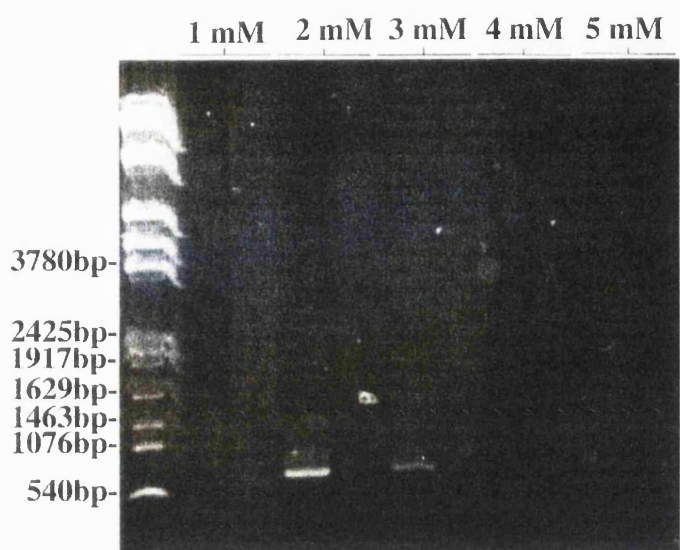
Figure 3.2 *Titration of [Mg²⁺]*

Reactions:

> 1 μl hi-titre 'phage stock
> 5 μl hi-temp forward primer (20 μM stock)
> 5 μl hi-temp reverse primer (20 μM stock)
> 1 μl dNTPs (20 μM stock)
> 10 μl 10x Vent buffer
> 2 μl 100 mM MgSO₄
> 0.5 μl Vent
>
> ddH_s O to 100 μl

Temperature cycling:

> 98 °C for 10' 1 cycle
>
> Add *Vent* DNA polymerase
>
> 98 °C for 1'
> 72 °C for 3' for 25 cycles
>
> 72 °C for 4'

Figure 3.3 *PCR conditions for the generation of ESTs*

70
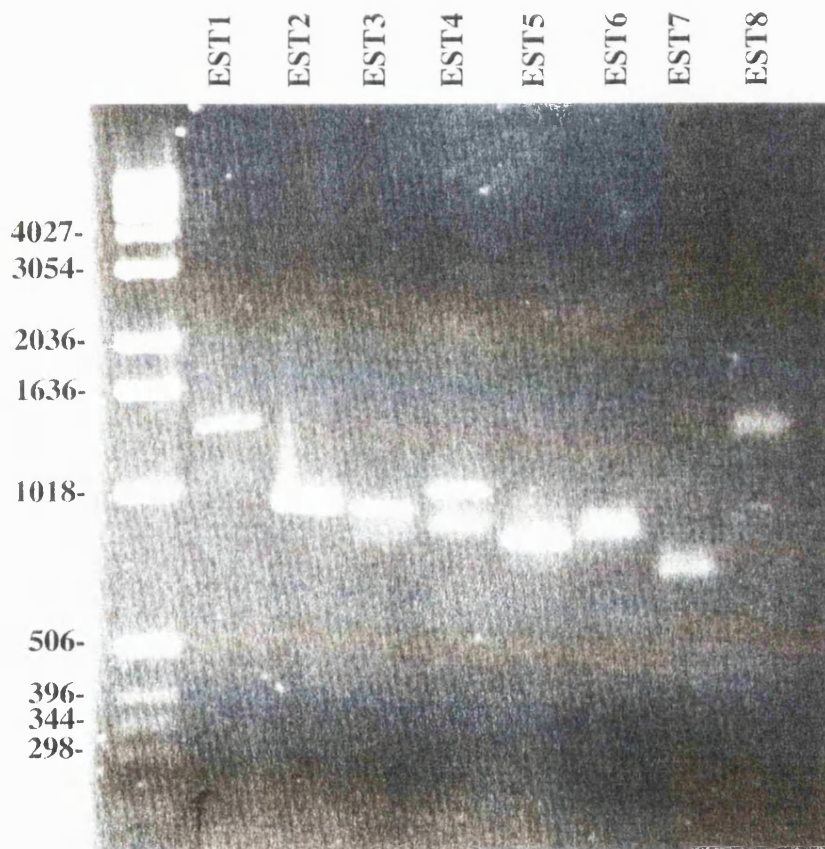
Figure 3.4   *PCR products generated from a* Chlamydomonas *cDNA library*

71

mM Tris-HCl pH 7.5 with a few drops of chloroform to prevent further bacterial growth. 10 µl of this plaque suspension was used as template for the PCR reaction. The cycling times and temperatures remained the same as those for the hi-titre PCR protocol.

3.2.4 Sequencing of ESTs

Sequencing primers were raised to the λgt10 arms inside the high-temperature PCR primers such that the PCR products could be sequenced using the Perkin-Elmer ABI Genetic Analyser 310 sequencer (figure 3.5). The PCR products were cleaned on a Microcon column (Amicon) prior to sequencing. The sequence data was obtained in the form of electropherograms (figure 3.6) and edited by eye to eliminate regions of vector, linkers and low quality data from the end of the sequencing run. With this type of sequencing accuracy starts to tail off at around 400 bases after which bases are frequently mis-called. Thus most of the ESTs generated were around 400 base-pairs long. Where necessary, accuracy was improved by human intervention, checking the automated base assignment against the trace data produced by the sequencer. The sequences were then converted to text files prior to similarity searching. A single sequence read from one end of each clone provides an identification tag that allows a match to be made against the public protein databases and known genomic sequences.

Sixteen clones with inserts of 600 base pairs or more were initially selected and sequenced from both ends. Ideally, the clones would be sequenced only from the 5' end since this is the coding region of the sequence. However, the library used for this pilot study is not uni-directional (i.e. clones may insert in either orientation) and a decision was made to sequence the pilot ESTs from both ends. After manual editing to remove regions of low quality data from the ends of each sequencing run, the sequences were converted into text files and compared against the NCBI databases using the BLAST (Basic Local Alignment Search Tool) interface available on the World Wide Web with the BEAUTY post processing (BLAST Enhanced Alignment Utility) available at the Baylor College of Medicine.

3.2.5 Nucleic acid sequence similarity searching

BLAST is the algorithm employed by sequence similarity searching programs utilising the statistical methods of Karlin and Altschul (Altschul, Gish et al. 1990). The BLAST programs first look for similar segments between the query sequence and a database sequence. The statistical significance of any matches found is calculated and matches that satisfy a threshold of significance are reported to the user. There are five different blast programs available that perform various tasks such as comparison of an amino acid sequence against a protein sequence database (blastp), comparison of a six-frame conceptual

λgt10 arms   λgt10 arms

forward   reverse
sequencing primer   sequencing primer
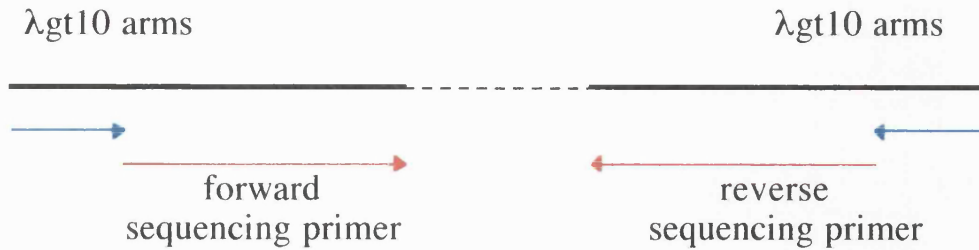
Random *Chlamydomonas* sequences or ESTs

```
TCCAAGCTCGAATTCCGGTTTTTTGAATCCTTGCCGTGNTACACAAC
AAGGGCAAATCGCAAACCGAACTGGCAANCAAAGGCCCAANGACCAT
TTTTATTCTTCTCAANGGGGCAGACATTACCCGCAAGCAAAACCCGA
ACACAAAACTGAACCTCCGTATCACACGTAACACCGCCACTTCGGTT
AATCGCACGTCCCTCGCGCTCCTAAGCCTGTGAAAAGANGCTCACAC
TGTATCGACACCGCGCTCCTCTTCATCTCCGCTCAATCACGCNAGTA
CATGCACTGCCGAGGGCCAAAAATTTACGCCGAGGGGGTGAATTGGT
GGCGTANGCNAACGCGTTCACGGCGGTGGGGTTGGCCANGTGGTCNT
CCAGGTCTGGATGGGGCCTTGCCG...
```

Compare with protein databases using BLASTX algorhithm

Sequence similarities to known proteins

e.g.

gi|295799 (X04472) ribulose bisphosphate carboxylase
[Chlamydomonas reinhardtii]

```
Query:     204 VQRPKSARDWQPANKRSV 151
               VQRPKSARDWQPANKRSV
Sbjct:     168 VQRPKSARDWQPANKRSV 185
```

Figure 3.5 *Sequencing strategy for the generation of ESTs*
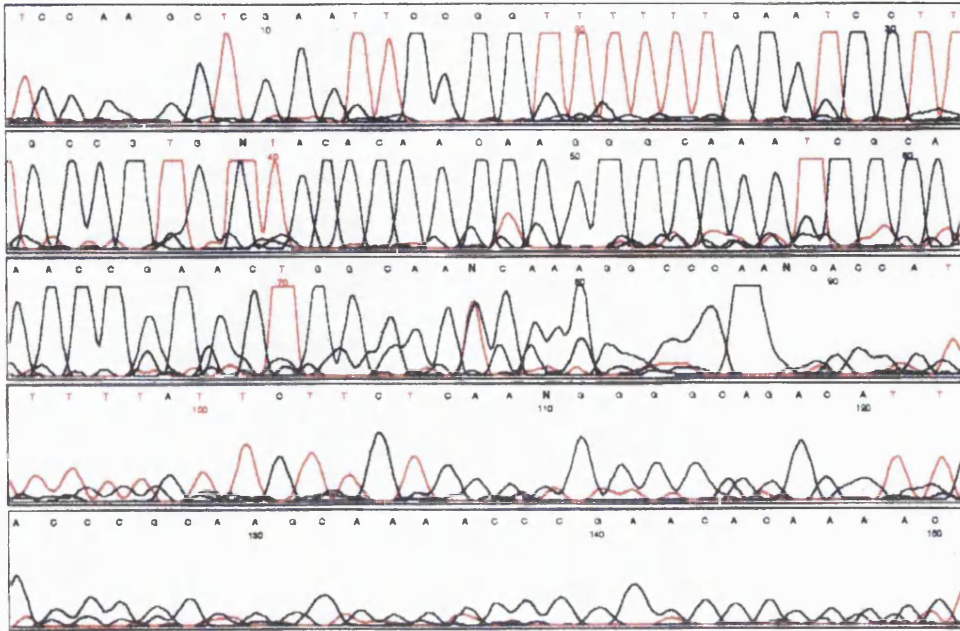
TCCAAGCTCGAATTCCGGTTTTTTGAATCCTTGCCGTGNTACACAACAAGGGCAAA
TCGCAAACCGAACTGGCAANCAAAGGCCCAANGACCATTTTTATTCTTCTCAANGG
GGCAGACATTACCCGCAAGCAAAACCCGAACACAAAACTGAACCTCCGTATCACAC
GTAACACCGCCACTTCGGTTAATCGCACGTCCCTCGCGCTCCTAAGCCTGTGAAAA
GANGCTCACACTGTATCGACACCGCGCTCCTCTTCATCTCCGCTCAATCACGCNAG
TACATGCACTGCCGAGGGCCAAAAATTTACGCCGAGGGGGTGAATTGGTGGCGTAN
GCNAACGCGTTCACGGCGGTGGGGTTGGCCANGTGGTCNTCCAGGTCTGGATGGGG
CCTTGCCGGTCCAATGGCTGGANAANAACCNACNTGNAAAAATACCNGNGGCGTCT
TATCTCTTCACTTCACCACNAAGTTNNGGTTCTCANCAAGCCAAGGGTCAAGCCCN
CGGGTTCAGGGTCAGGCCCCCCANGGGCCCNTTGGNCGTTGCNCNCNGCCTCCNAA
CTNNAAGCNGTTNAAAAGANNGNTTANNGTTCTTCAANTTTTCGCCCCCCNAAATT
TGGNCCGNTNAACAAACCNCNANNTTCTTTTNCNNNNGGTTNNNCCAACCNTTNNN
GGNTNNNCCCCNNTNTNNTNCTNCNAAANNTTNCNTNCNNANCTTTNGGTTCCNAA
NNAN

Figure 3.6 *Example of electropherogram sequence. This sequence, obtained for EST 9, exhibits a poly-(A) tail (represented as a row of six thymine bases) next to the EcoRI linker site at the beginning of the sequence (shown in blue).*

74

translation of the query sequence against the NCBI non-redundant protein sequence database (blastx), comparison of a nucleotide query sequence against a nucleotide sequence database (blastn), comparison of a protein sequence against a nucleotide sequence database translated in all six reading frames (tblastn) and comparison of six-frame nucleotide sequence translations against the six-frame translations of a nucleotide sequence database (tblastx). For the purposes of this project the program blastx was used. This program allows for degeneracy in the genetic code and provides useful output for this type of project.

The BLAST algorithm

The fundamental unit of the BLAST algorithm is the High Scoring Segment Pair (HSP). The task of finding HSPs begins with identifying short words of length W in the query sequence that either match or satisfy a threshold score, T, when aligned with a word of the same length in the database sequence. T is referred to as the neighbourhood word score threshold. These initial neighbourhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence as far as the cumulative alignment score can be increased and halted when the cumulative word score falls off by a quantity X from its maximum achieved value. The parameter E establishes a statistical significance threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP within the database search and may be thought of as the number of matches one expects to observe by chance alone. Any database sequence whose matching satisfies E is reported in the program output. These parameters may be adjusted by the user if required. For example, if sensitivity is not an issue but rapidity of the search is important, the parameters may be adjusted to achieve higher speed at the expense of accuracy (e.g. increase W by one and T by 10-50%). Default parameters for the BLAST programs have been chosen to give moderate sensitivity in detecting matches and these default parameters were found to give satisfactory results for this project.

The BLAST output format is as follows: first comes an introduction to the program, then a series of one line descriptions of matching database sequences followed by the actual sequence alignments. Finally the parameters and other statistics gathered by the search are listed. The one-line sequence descriptions and summaries of results are useful for identifying biologically interesting database searches and correlating this interest with the statistical significance estimates. The database sequences are sorted by increasing P-value (probability). Several statistics are used to describe each HSP: the raw alignment score; the raw score converted to bits of information by multiplying by a factor lambda; the number of times one might expect to see such a match by chance; the probability of observing such a match and the number and fraction of residues for which the alignment scores have positive values (figure 3.7). In general, the lower the probability value, the greater the significance.

```
BLASTX 1.4.11 [24-Nov-97] [Build 24-Nov-97]
Query= tmpseq_1
       (490 letters)
Translating both strands of query sequence in all 6 reading frames

Database:  Non-redundant GenBank CDS
           translations+PDB+SwissProt+SPupdate+PIR
           277,055 sequences; 82,859,678 total letters.


                                                          Smallest
                                                            Sum
                                         Reading  High  Probability
Sequences producing High-scoring Segment Pairs:   Frame Score  P(N)        N

gi|2183271 (AF002133) MAV305 [Mycobacterium avium]    -3    61  0.9996      1


gi|2183271 (AF002133) MAV305 [Mycobacterium avium]
           Length = 305

  Minus Strand HSPs:

 Score = 61 (27.9 bits), Expect = 8.0, P = 1.0
 Identities = 12/23 (52%), Positives = 12/23 (52%), Frame = -3

Query:   278 RATWRPNSGRLKIRGCAGFARLP 210
             R  WRPN   K R CAG A  P
Sbjct:   265 RLLWRPNDNASKTRRCAGLAGEP 287

Parameters:

  V=100
  B=50
  H=0
  -filter
  SEG

  Lambda     K        H
    0.318   0.135    0.401

  Cutoff to enter 2nd pass: >= 46 ( 0.0 bits)

  E     S     T1     T2     X1     X2     W      Gap
  10.0        62     12     12     -16    -22    40      50
```

Figure 3.7 *Example of BLASTX program output*

However, even if a reported probability is low, it is necessary to look at the sequences involved in the alignments to see if there is something remarkable about them such as their residue composition which would tend to produce lots of matches. When a sequence contains one or more regions of biased residue composition, there may be uninteresting database hits reported first due to their low probability scores. The 'filter' option may be used to avoid these uninteresting hits and get to the interesting ones sooner, but biological knowledge should be used to temper such conclusions, and for the purposes of this project it was decided to use the default parameters set by the BLAST program and to edit the output by eye so as not to miss interesting matches.

In summary, BLAST is an approximate word match program that produces a list of the sequence identifiers (i.e. locus name and accession numbers) and the title lines of statistically significant matches, followed by a display of the alignments of the query with each of the matched sequences. For a sequence database search result to be informative, two criteria must be met. Firstly the query sequence must have a statistically significant match to the database greater than that expected by chance alone. Secondly there must be information available about the function of the sequence matched. It is quite common that the functions of matched sequences are not obvious from the search results. Sequence titles are often uninformative and one must laboriously retrieve and scan the full sequence database reports to look for annotations that may identify the biological function of the matched sequence. In addition, functionally important domains such as enzyme active sites are not noted as such in sequence database records. To allow easier identification of the functions of sequences matched in our database searches, we have employed BEAUTY, an enhanced version of the NCBI BLAST search tool (Worley, Wiese et al. 1995).

BEAUTY post processing

BEAUTY provides information on sequence family membership, the relative locations of conserved regions and the locations of annotated domains and sites within the query sequence. This information is incorporated directly into the BLAST search results. These enhancements make it much easier to identify the functions of matched sequences, which is particularly important when trying to analyse the biological significance of weak database hits. The BEAUTY Post-Processor is available at the Baylor College of Medicine Website. New databases of conserved regions and of functional domains for protein sequences have been created for use with BEAUTY. A Conserved Regions Database, containing the locations of conserved regions within protein sequences has been constructed by clustering the Entrez [*] protein database into families, aligning each family using a multiple sequence alignment program and scanning the multiple alignments to locate the conserved regions within each alignment. A separate Annotated Domains Database has been constructed by extracting the locations of all annotated domains and sites from sequences represented in

```
Locally-aligned regions (HSPs) with respect to query sequence:

Locus_ID                        Frame 2 Hits
gi|2117963|pir||I5101    |  _____
gi|129415|sp|P04056|P    |        _____        ____
gi|129455|sp|P00628|P    |        _____
gi|129471|sp|P20256|P    |        _____        ___
gi|2431864               |                   ___  _____
gi|129511|sp|P00594|P    |        _____
gi|1362592|pir||S5742    |  ____
gi|464331|sp|P00598|P    |  _____
gi|804798                |  _____
gi|1703010|sp|P50406|    |                   _____
gi|804794                |  _____
gi|804796                |  _____
gi|129497|sp|P14411|P    |        _____
gi|346607|pir||JN0480    |        _____
gi|129512|sp|P19000|P    |        _____
Prosite Hits:                     __

Query sequence:          |                |            |            |
144
                         0               50          100
```

Prosite hits:
    PA2_HIS   Phospholipase A2 histidine active site. 45..52

Figure 3.8 *Example of BEAUTY program output. A figure is added showing the relative location of each hit (HSP) within the query sequence with the accession number linked to the individual reports. In addition the query sequence is matched against the PROSITE pattern database and the location of all pattern matches within the query sequence is displayed. Multiple hits within the same region of a query sequence may indicate a functionally important site (e.g. an ATP-binding domain).*

```
Local hits (HSPs):                      _____         _____
Annotated Domains:             _____  _____  _____  _____

Database sequence:             |                 |                 |           |
118
                               0                 50               100


─────────────────

Annotated Domains:
    Entrez            active site                                   48
    Entrez            Calcium binding region                        49
    PRINTS            PHPHLIPASEA21: Phospholipase A2 motif I     2..12
    PRINTS            PHPHLIPASEA22: Phospholipase A2 motif II   18..36
    PRINTS            PHPHLIPASEA25: Phospholipase A2 motif V    84..100
    BLOCKS            PA2_HIS: Phospholipase A2 histidine prot   1..13
    BLOCKS            PA2_HIS: Phospholipase A2 histidine prot   25..57
    BLOCKS            PA2_HIS: Phospholipase A2 histidine prot   84..99
    PROSITE           PA2_ASP: Phospholipase A2 aspartic acid    88..98
    PROSITE           PA2_HIS: Phospholipase A2 histidine acti   44..51

─────────────────
```

Figure 3.9 *A figure is added for each BLAST hit showing:*
a) the positions of the local hits (HSPs) and
b) the location of any annotated domains and sites within each
matched sequence.
Hits matching known domains and sites are readily discernible.

Entrez, Prosite, Blocks and Prints protein motif databases. BEAUTY first performs a BLAST search of those Entrez sequences with conserved regions and/or annotated domains. The information from the Conserved Regions and Annotated Domains databases is then used to create a schematic display for each matched sequence that allows the direct comparison of the relative locations of the conserved regions, the annotated domains and sites and the locally aligned regions matched in the BLAST search. A figure is added to the BLAST search results summarising information on the locations of local hits and of any annotated domains and sites. The relative location of each hit (HSP) within the query sequence is shown with the accession number linked to the individual reports listed below (figure 3.8). A second figure showing the locations of any annotated domains or sites within each matched sequence is added for each BLAST hit (figure 3.9). Hits matching known domains and sites are therefore readily discernible. In addition, the query sequence is matched against the PROSITE pattern database, and the locations of all pattern matches within the query sequence are displayed. Multiple hits within the same region of a query sequence may indicate a functionally important domain. In summary, by incorporating annotated domain and site information directly into BLAST search results, BEAUTY can greatly improve the identification of weak but functionally significant matches in database searches. As a result the time needed for a scientist to fully evaluate the BEAUTY search results is significantly less than the time needed to evaluate a comparable BLAST search result.

The sixteen sequenced clones were divided into four groups based on the match of ESTs from both cDNA ends to the public sequence databases as shown in Table 3 a. A match was declared significant even if there were hits for the EST from only one of the extremities (Table 3 b).

3.2.6 Previously identified *Chlamydomonas* genes

Those clones that exhibited exact homology with genes previously characterised from *Chlamydomonas* correspond to two highly expressed *Chlamydomonas* genes: those for ferredoxin and the small unit of ribulose-1,5-bisphosphate carboxylase.

The [2Fe-2S] containing protein, ferredoxin, functions as the central molecule for distributing electrons originating from the oxidation of water during oxygenic photosynthesis. The electrons are passed to a number of ferredoxin-dependent enzymes in the chloroplasts of higher plants as well as in algae and cyanobacteria. The proteins contain, per molecule of protein, two non-haem irons and two inorganic sulphides arranged in a single cluster. The region in the vicinity of the cluster is the most highly conserved portion of ferredoxin, with regions of the protein remote from the cluster showing considerably more variability (Knaff and Hirasawa 1991). In higher plants and algae the gene coding for

80

EST number	Class
10, 12	Clones that exhibit exact homology with genes previously characterised from *Chlamydomonas* e.g. ferredoxin, rubisco
9, 13, 14, 15	Putative identification based on similarities to known *Chlamydomonas* multigenes i.e. cab II, hydroxyproline glycoprotein precursor.
3, 4, 5	Putative identification based on similarities to non-*Chlamydomonas* genes e.g. cytochrome p450, negatively phytochrome regulated protein.
1, 2, 6, 8, 11, 16	Clones that gave no match to database sequences. These genes can be assumed to represent new *Chlamydomonas* genes, but cannot be identified. In some cases a protein motif may be identified giving a clue to the function of the protein e.g. one clone exhibited a leucine zipper.

Table 3 a   *Table of classes of sixteen original ESTs*

EST	Size	Database matches?	Protein motifs?
1	1000	no convincing matches	zinc finger
2	1200	no matches	
3	650	λ transcription repressor protein	zinc finger
4	700	retinoic acid receptor beta-3 cytochrome P450 (mouse)	leucine zipper
5	1200	phospholipase A2 receptor negatively phytochrome regulated protein (duckweed)	calcium binding motif phospholipase motif
6	1200	no convincing matches	
7	1375	no convincing matches	leucine zipper prokaryotic membrane lipoprotein
8	1375	no convincing matches	
9	1600	chlorophyll a-b binding protein of LHC	
10	1250	*Chlamydomonas reinhardtii* ferredoxin	metal-binding site: iron-sulphur (2Fe-2S)
11	1151	no convincing matches	DNA-binding site: homeobox prokar_lipoprotein: Prokaryotic membrane
12	1280	Chlamydomonas RbcS2	atp/gtp-binding site motif a (p-loop). rubisco small 1: rubisco small subunit rubisco small 2: rubisco small subunit
13	1050	hydroxyproline-rich glycoprotein precursor (*Chlamydomonas*)	Domain: poly-GLN
14	1100	chlorophyll a/b-binding protein - *Chlamydomonas reinhardtii* microtubule associated protein	MAP2/tau repeat homology
15	950	glycoprotein precursor from a variety of organisms	
16	1600	no convincing matches	

Table 3. b   *Summary of BLAST search results for ESTs 1-16*

ferredoxin is a nuclear gene (*PetF*) and ferredoxin is synthesised in the form of a precursor protein with an amino-terminal sequence that is required for targeting the precursor for import across the chloroplast membrane into the stromal space. The signal sequence is subsequently cleaved by a specific protease to generate the mature peptide. The insertion of the [2Fe-2S] probably takes place in the stroma. The strongest database matches (i.e. the highest probability matches) were obtained for our ferredoxin cDNA to the metal binding site of ferredoxin for both *Chlamydomonas* ferredoxin precursor (accession gi|1009714 U29516) and to the *Chlamydomonas* ferredoxin protein itself (accession gi|65759|pir). (See figure 3.10). Strong matches were observed for the clone in both orientations. Significant matches were also shown to 50 ferredoxins from other species, with a further 93 matches going unreported due to a cut-off value of 50 being applied by the program. These matches represent functionally conserved regions of the protein.

A similar situation was observed for a second of the pilot ESTs which was found to show an exact database match to a highly expressed nuclear gene *RBCS2*, that encodes the small unit of ribulose-1,5-bisphosphate carboxylase or rubisco. Ribulose-1,5-bisphosphate carboxylase/oxygenase is the key photosynthetic enzyme that catalyses the first step of CO_2 fixation. This enzyme is localised to the chloroplast and comprises eight nuclear-encoded small subunits and eight chloroplast encoded large subunits. In green plants the small subunits are coded by a family of two or more *RBCS* nuclear genes, synthesised as 20 kDa precursors in the cytoplasm, and processed to 15 kDa during transport into the chloroplast. In *Chlamydomonas* the small subunit is encoded by two genes *RBCS1* and *RBCS2* (Goldschmidt Clermont 1986) the gene products of which differ by four amino acids and are differentially expressed during plant development (Khrebtukova and Spreizer 1996). The highest probability match was to *Chlamydomonas RBCS2* and *Chlamydomonas* rubisco small-unit precursor which showed near perfect sequence homology (figure 3.11). A host of matches to rubisco from other species followed, showing matches to the same highly conserved region. Descriptions of 178 database sequences were not reported due to the limiting value of parameter V=50.

Even in such a small EST sample, two of the most highly expressed *Chlamydomonas* genes are represented. Evidence from other EST analysis projects shows that highly expressed genes form a significant percentage of clones for most projects. In order to avoid continually sequencing highly expressed clones some groups have initiated sequencing projects that utilise a sorted DNA library, where highly expressed clones are removed by hybridisation techniques (Hoog 1991).

3.2.7 A putatively identified *Chlamydomonas* multigene

The second class of ESTs is that representing *Chlamydomonas* genes that have been putatively identified based on partial similarity to previously sequenced *Chlamydomonas*

Forward Strand Matches:

```
gi|65759|pir||FEKM ferredoxin [2Fe-2S] - Chlamydomonas reinhardtii
          Length = 94
```

```
Frame -1 hits (HSPs):                        _____        _____
Frame -3 hits (HSPs):                                 _____

Database sequence:     |         |         |         |         |         |
94
                       0        20        40        60        80
```

```
Query:   345 QVAAGTVDQSDQSFLDDAQMGNGFVLTCVAYP 250
             +VAAGTVDQSDQSFLDDAQMGNGFVLTCVAYP
Sbjct:    48 KVAAGTVDQSDQSFLDDAQMGNGFVLTCVAYP 79


Query:   251 PTSDCTIQTHQEEALY 204
             PTSDCTIQTHQEEALY
Sbjct:    79 PTSDCTIQTHQEEALY 94


Query:   380 SCRAGACSKCAGK 342
             SCRAGACS CAGK
Sbjct:    36 SCRAGACSSCAGK 48
```

Reverse Strand Matches:

```
gi|65759|pir||FEKM ferredoxin [2Fe-2S] - Chlamydomonas reinhardtii
          Length = 94
```

```
Frame  3 hits (HSPs):    _____        _____
Frame  2 hits (HSPs):             _____                 _____

Database sequence:     |         |         |         |         |         |
94
                       0        20        40        60        80
```

```
Query:   168 LDXPYSXXAGACSXCAGKVAAGTVDQSDQSFMDDAXMGNGFVLTCMAYP 314
             LD PYS  AGACS CAGKVAAGTVDQSDQSF+DDA MGNGFVLTC+AYP
Sbjct:    31 LDLPYSCRAGACSSCAGKVAAGTVDQSDQSFLDDAQMGNGFVLTCVAYP 79


Query:   113 TIECPADTYILDAAXEAGL 169
             TIECPADTYILDAA EAGL
Sbjct:    13 TIECPADTYILDAAEEAGL 31


Query:   311 PTSDCTXQT 337
             PTSDCT QT
Sbjct:    79 PTSDCTIQT 87


Query:    75 YKVTLEDPFG 104
             YKVTL+ P G
Sbjct:     1 YKVTLKTPSG 10
```

Figure 3.10 *Figure showing database matches to* Chlamydomonas reinhardtii *ferredoxin.*

Forward Strand Matches:

```
gi|762923 (X04471) ribulose bisphosphate carboxylase [Chlamydomonas
             reinhardtii]
             Length = 140


Frame -1 hits (HSPs):                                    _____
Frame -2 hits (HSPs):                          _____
Frame -3 hits (HSPs):                  _____


Database sequence:      |              |              |               | 140
                        0             50            100
```

```
Query:    289  QAFPDAYVXLVAFDNQKQVQIMGFLVQRPKSARDWQPANKRSV 161
               +AFPDAYV LVAFDNQKQVQIMGFLVQRPK+ARD+QPANKRSV
Sbjct:     98  KAFPDAYVRLVAFDNQKQVQIMGFLVQRPKTARDFQPANKRSV 140


Query:    393  VSCLXXDNRXWTMWKLPMFGCXDPMQVLREIVACTKLSP 277
               VSCL  DNR WTMWKLPMFGC DPMQVLREIVACTK  P
Sbjct:     63  VSCLYYDNRYWTMWKLPMFGCRDPMQVLREIVACTKAFP 101


Query:    434  KAXVSNESAIRF 399
               KA VSNESAIRF
Sbjct:     49  KAYVSNESAIRF 60
```
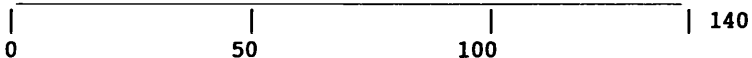
Reverse Strand Matches:

```
gi|762923 (X04471) ribulose bisphosphate carboxylase [Chlamydomonas
             reinhardtii]
             Length = 140


Frame  1 hits (HSPs):      _____


Database sequence:      |              |              |               | 140
                        0             50            100
```

```
Query:     55  MMVWTPVNNKMFETFSYLPPLSDEQIAAQVDYIVANGWIPXLEFAESDKAYVSNESAIRF 234
               MMVWTPVNNKMFETFSYLPPL+DEQIAAQVDYIVANGWIP LEFAE+DKAYVSNESAIRF
Sbjct:      1  MMVWTPVNNKMFETFSYLPPLTDEQIAAQVDYIVANGWIPCLEFAEADKAYVSNESAIRF 60


Query:    235  GSVSCLYXDDRXWTCGXCPXSAAATPCXLLRXIVAXTKAFPDAYVRLVACDNQKXVQIXG 414
               GSVSCLY D+R WT   P      P +LR IVA TKAFPDAYVRLVA DNQK VQI G
Sbjct:     61  GSVSCLYYDNRYWTMWKLPMFGCRDPMQVLREIVACTKAFPDAYVRLVAFDNQKQVQIMG 120
```

Figure 3.11 *Figure showing database matches to* Chlamydomonas
reinhardtii *Rubisco*

genes. These genes are thought to represent previously uncharacterised members of multigene families, hence their similarities to known *Chlamydomonas* proteins. For example, EST 13 showed homology to hydroxyproline-rich glycoprotein precursor, a constituent of the *Chlamydomonas* cell wall. The *Chlamydomonas* cell wall is a complex, multi-layered structure that contains 20-25 proteins and glycoproteins, many of which are enriched in hydroxyproline. Unlike the cell walls of higher plants it has little if any cellulose (Iman and Snell 1988). The similarities of the EST to gi|2384728 (AF015883) the hydroxyproline-rich glycoprotein gas28p precursor from *Chlamydomonas,* suggest that the DNA encodes a glycoprotein precursor. However, the lack of complete homology between the two as in the case of ferredoxin and rubisco suggest that this cDNA has not been previously sequenced but is a member of the same gene family.

A second EST (EST 9) showed a high degree of similarity to *Chlamydomonas* chlorophyll a/b binding protein from photosystem II. This EST is a putative new member for the *Chlamydomonas* family of CAB proteins although further analysis is required to verify that the sequence represents a novel CAB protein from *Chlamydomonas*. Photosystems I and II (PSI and PSII) contain different but structurally related chlorophyll a/b binding proteins. These proteins, with their associated pigments, are known as the light harvesting complexes (LHCs). The LHCs capture light energy and funnel it into the photosynthetic pathway. Two cab genes have so far been cloned and characterised from *Chlamydomonas*, the *Lhca1*1* gene encoding a 20 kDa LHCI polypeptide associated with photosystem I (Hwang and Herrin 1993) and the *cabII-I* gene which encodes a 28 kDa LHCII polypeptide associated with photosystem II (Imbault, Wittemer et al. 1988). The CAB proteins are encoded in the nuclear genome where they form a multigene family, thought to consist of 3-7 members (Bassi and Wollman 1991; Bassi, Soen et al. 1992). The proteins are translated in the cytoplasm and then imported into the chloroplast, where the transit peptide is cleaved and the protein inserted into the thylakoid membrane. CAB polypeptides show overall amino acid sequence homology in pairwise comparisons that range from 99% to 30%. These *Chlamydomonas Lhc1*1* and *cabII-I* polypeptides show approximately 50 % and 75 % overall identity to their higher plant counterparts. The CAB proteins have three hydrophobic regions representing transmembrane helices which are well conserved for most species. Of these, transmembrane helices I and III are the most strongly conserved sequences among all CAB polypeptides. Based on this description, a third class of CAB protein *L1818* has been identified in *Chlamydomonas. L1818* shows homology to *Lhca1*1*, *cabII-I* and all known CAB genes in the transmembrane regions, is light-induced and part of a multigene family. However its overall similarity to all CAB genes is less than 30%. Nevertheless, *L1818* polypeptides are likely to be relatives of CAB polypeptides (Savard, Richard et al. 1996). The translated EST 9 shows similarity but not complete homology to both the
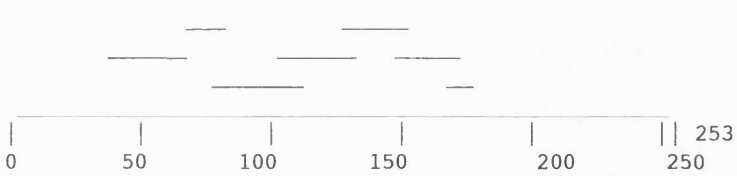
gi|81235|pir||JS0075 chlorophyll a/b-binding protein - *Chlamydomonas reinhardtii*

Length = 253

```
Frame  3 hits (HSPs):                        __        _____
Frame  2 hits (HSPs):              _____        _____    ___
Frame  1 hits (HSPs):                   _____        __

Database sequence:    |      |      |      |      |     || 253
                      0      50     100    150    200   250
```

Plus Strand HSPs:

```
Query:  118 TFKRYRELXLIHARWAMLGALGCITPELLAK 210
            TFKRYREL LIHARWAMLGALGC+TPELLAK
Sbjct:   78 TFKRYRELELIHARWAMLGALGCLTPELLAK 108


Query:  194 PSCWPRNGIPFGEAVWFKAGAHIFAXGGLE 283
            P    ++G   FGEAVWFKAGA IF+ GGL+
Sbjct:  103 PELLAKSGTKFGEAVWFKAGAQIFSEGGLD 132


Query:    2 DRPXFLGPFSEGDTPAYLTGXSXGXXG 82
            +R  +LGP+SE  TPAYLTG   G  G
Sbjct:   40 NRAKWLGPYSENATPAYLTGEFPGDYG 66


Query:  279 LNYLGNXNLIHAQSIIAT 332
            L+YLGN +L+HAQ+I+AT
Sbjct:  131 LDYLGNPSLVHAQNIVAT 148


Query:   84 WXTAGLSADPD 116
            W TAGLSADP+
Sbjct:   67 WDTAGLSADPE 77


Query:  335 AFXVVVMGPG*XYRANGGPWGK 400
            A  V++MG    YR NGGP G+
Sbjct:  150 AVQVILMGLIEGYRVNGGPAGE 171


Query:  388 PLGQGLDP 411
            P G+GLDP
Sbjct:  168 PAGEGLDP 175
```

Figure 3.12 *Protein sequence comparison between EST 9 and* Chlamydomonas
reinhardtii *CABII-I. The residues highlighted in green indicate
regions of homology between the two sequences, those highlighted
in blue represent conserved changes in the amino acid sequence and
those highlighted in red indicate regions where there is no homology
between the two sequences.*

87

Chlamydomonas CAB genes and to CAB genes from other plant systems (figure 3.12), and thus represents a putative new member for the CAB

protein family. It is possible that the differences in protein sequences could arise from a high degree of sequencing error possible in this type of project, where cDNAs are amplified by PCR and subject to single pass sequencing. Certainly, some degree of sequencing error is represented by the translated protein as the CAB sequence was found to slip out of frame on several occasions.

This putative identification of a new member of the *Chlamydomonas* CAB gene family by an EST approach demonstrates that EST analysis is effective as a means of identifying novel *Chlamydomonas* genes. The placement of new genes on the *Chlamydomonas* genetic map will provide new genetic markers for the genome. ESTs mapped to their positions on a physical map of an organism become known as Sequence Tagged Sites (STSs) and are valuable genetic markers. In order to completely characterise the new CAB protein it will be necessary to pull out a full length clone of the CAB gene and to sequence it in its entirety. As a complete indexed cosmid library exists for *Chlamydomonas* (Zhang, Herman et al. 1994) this process may be achieved with relative ease. However, the process from EST to full-length genomic sequence is not straightforward, so it was decided not to pursue this avenue since the object of the EST pilot project had been achieved. One foreseeable problem in the cloning of the full length CAB gene is that of difficulties imposed by cross-hybridisation among members of the same gene family. It is also important to bear in mind that the full-length cDNA represents only the sequence of the mRNA and that the genomic sequence will contain introns and other flanking regions. It may also be difficult to clone the full sequence of mRNAs that are large in structure or could include incorrectly spliced messages. G+C rich sequences easily form secondary structure and it can be difficult to obtain sequence data for these regions. The effort involved may range from a matter of weeks in the case of a short easily cloned gene to more than a year. It may also be necessary to screen multiple cDNA libraries or to use various cloning-based or PCR-based approaches.

3.2.8 Putative identification of ESTs based on similarities to non-*Chlamydomonas* genes

Two clones showed significant sequence similarity to non-*Chlamydomonas* genes. For example, one clone (EST 5) showed significant sequence similarity to negetively regulated phytochrome protein from duckweed. Such results must be treated with caution, and detailed further analysis is required to achieve positive identification. The existence of a library of such clones is useful however, and could be used as a first-line approach in looking for genes encoding novel proteins from the organism.

	EST number	P(N) *	Putative Identification
Previously identified *Chlamydomonas* **genes**	36	8.2e-28	ADP-ribosylation factor
	81	8.7e-38	Ferredoxin
	83	2.1e-55	RBCS2
	94	0.96	Chlorophyll a/b binding protein
Provisional Identification based upon sequence similarity to other known proteins	34	3.6e-19	CAB II (*Arabidopsis*)
	37	5.0e-50	Lambda transcription factor
	43	4.2e-19	Soluble starch synthase (bacterial)
	44	0.98	CAB II (*Pyrpbotrys stellata*)
	45	0.83	Cytochrome P450 (mouse)
	46	5.9e-18	Phytochrome protein (duckweed)
	60	5.0e-14	PsaL from a variety of plant *Sp.*
	67	5.0e-14	"
	70	5.9e-18	"
	72	8.3e-17	"
	77	5.2e-17	"
	74	0.74	CAB II form a variety of plant species
	75	1.4e-64	60S acidic ribosomal protein P2 (*Trypanosome cruzi*)
	68	0.14	60S ribosomal protein (*Arabidopsis*)
	84	0.42	Hydroxyproline-rich glycoprotein precursor
	95	—	"
Unidentified genes	4, 6, 9, 10, 11, 16, 32, 33, 35, 40, 42, 49, 57, 71, 73, 79, 82, 85, 28, 48, 80		
Clones with no insert	41 in total		
Sequences that gave no match to any known protein	18 in total		

Table 3 c *Table of results from EST analysis of 100 independent* Chlamydomonas cDNA clones.

*Where P(N) is the probability of observing such a match. In general, the lower the P value, the greater the significance (see page 75).

3.2.9 Sequence analysis and similarity searches of a further 84 ESTs

Preliminary EST analysis of the remaining 84 clones produced by the EST PCR protocol further demonstrated the effectiveness of the procedure and produced some interesting results. In total, 100 clones were sequenced and were found to include the genes encoding previously characterised *Chlamydomonas* proteins rubisco, cabII, ferredoxin and a previously characterised ADP-ribosylation factor (Memon et al 1995). The BLAST searches also indicated the presence of a number of other interesting clones in the library including a soluble starch synthase. In addition, several clones encoding a sequence that is remarkably similar to *psaL*, a gene encoding subunit V of photosytem I from a number of organisms which has not so far been cloned from *Chlamydomonas*. Also included were a ribosomal 60S protein that showed high homology to the 60S from *Arabidopsis* and yeast, and another 60S acidic RNA protein that showed similarity to the 60S acidic ribosomal protein P2 from *Trypanosomo cruzi* (Table 3 c).

3.3 Discussion

3.3.1 Evaluation of the EST protocol

The object of the project has been achieved - that is the development of an efficient and rapid procedure for the isolation of ESTs from *Chlamydomonas*. One limitation of single pass sequencing of PCR products is that the sequence generated may have a high degree of error. However, for tag sequencing great accuracy is not required and provided significant similarities in the database can be detected, the objective of the screen will be achieved. These sequences in combination with their positioning on the *Chlamydomonas* genetic map will not only constitute a new set of molecular markers for genome analysis but also provide a direct route for the *in vivo* analysis of their gene products.

Some enhancements are desirable for the EST generation procedure. One improvement would be a means of eliminating redundant sequences from the 3' untranslated region sequences by implementing a directionally cloned library in order to preferentially target the coding regions of the cDNAs. For sequence comparisons, data from the 5' end of a poly(A)-selected cDNA should be much more informative than data from the untranslated 3' tail and sequences localised next to the 3' end of the cDNA will account for a proportion of the redundant clones in a non-directional library. Such a library, constructed in a λ Uni-ZAP vector has been obtained by the Purton laboratory and will be used in the next stages of the EST sequencing project.

A second improvement to the process may be achieved through the use of a cDNA library that has been screened through hybridisation to remove clones coding for highly expressed genes that may be represented several times within a library. As the aim of any EST project is to isolate novel, previously uncharacterised genes from the tissue under investigation it

has been found that the use of sorted DNA libraries can be useful in some cases as a means of reducing the number of previously identified clones in a library and to avoid repetitive sequencing of the same highly expressed clones. Hoog compared two different approaches for their effectiveness in isolating novel genes in a mouse EST sequencing project: random versus differential complementary cDNA cloning methods (Hoog 1991). In the differential approach, only the cDNA clones containing rare sequences (as determined by preliminary clone hybridisation) are further analysed; in the random approach, cDNA clones were isolated at random from the cDNA library. More than two hundred cDNA clones altogether were analysed, using a PCR mediated amplification and sequencing strategy. A comparison of these sequences to nucleic acid and protein sequence databases revealed that 84 % of the isolated rare cDNA clones represented novel genes. In contrast, less than 63 % of the cDNA clones isolated at random from cDNA libraries were novel sequences. Thus the probability of isolating new, previously uncharacterised genes from cDNA libraries can be markedly improved by focusing efforts on clones containing rare sequences. However, this approach also precludes picking up clones of abundant protein isoforms or multigenes such as the CAB proteins or the hydroxyproline-rich glycoprotein precursor molecules, as found in a random clone sequencing approach. These gene families show similar translated amino acid sequences in their open reading frame regions but may differ in their untranslated regions and in areas of the proteins that are not well conserved. Therefore gene families cannot be identified in this way.

3.3.2 The *Chlamydomonas* genome sequencing project

Sequencing the entire genome is an expensive and technically difficult effort which will only be achieved through the collaboration of the *Chlamydomonas* community. Given the size of the genome (~100 Mb), its high G+C content and the presence of numerous repeats, sequencing of the entire genome is a long-term goal. In the short term, proposals for managing such a project include:

i) The development of an Expressed Sequence Tag database to identify genes expressed in *Chlamydomonas*. The cDNAs sequenced could be isolated from cells grown under different conditions. These ESTs could be used as molecular markers in the building up of a physical map of the genome, and provide information on the expressed gene content of cells grown under varying conditions.

ii) The generation of a contiguous map of *Chlamydomonas* BAC (Bacterial Artificial Chromosome) clones and the linking of the physical and genetic maps, facilitated by the molecular mapping efforts.

iii) The intron/exon borders of genomic DNA should be determined if genomic sequence is to be of use. cDNAs that are already available along with their corresponding genomic clones should be used in the design of algorithms that can learn to recognise intron/exon borders.

iv) A collection of mutants maintained in manageable size lots (e.g. 100) generated by insertional mutagenesis. These mutant pools could be made available for researchers to screen by PCR for mutations in genes of interest. This protocol is dependent on the ability to routinely amplify sequence from a disrupted gene using primers specific to the gene of interest and to the insertional mutagen when the genomic DNA is diluted from 99 other transformants. The process is also dependent on whether freezing protocols are efficient enough to allow the freezing away of pools of 100 transformants. Maintenance of individual transformants on agar would be extremely labour intensive and maintenance of the stocks as pools on agar   may lead to selection against certain insertional mutants within the population.

v) The development of cDNA microarrays for hybridisation to look at the pattern of many genes simultaneously and to evaluate mutants that may be deficient in regulating these patterns.

The pilot study of EST analysis in *Chlamydomonas* has produced a practical and robust method for the generation and analysis of ESTs that, given the current interest in establishment of a *Chlamydomonas* genome sequencing project, should prove invaluable to genome researchers.

Chapter 4

GREEN FLUORESCENT PROTEIN - A SUITABLE REPORTER
FOR *CHLAMYDOMONAS REINHARDTII?*

93

Chapter 4:

Green fluorescent protein - a suitable reporter for *Chlamydomonas reinhardtii?*

4.1 Introduction

The development of DNA transformation techniques for the nuclear and chloroplast genomes of *Chlamydomonas* has greatly expanded the usefulness of this organism, allowing molecular dissection of complex cellular processes. Nuclear transformation technology has opened the door to gene expression and protein localisation studies *in vivo* using reporter genes and other molecular tools. Two *Chlamydomonas* genes have been employed as reporters of promoter activity. Firstly, by fusing a promoter to the coding region of the *ARS* gene (encoding periplasmic arylsulphatase) it is possible to monitor promoter activity in the transformants by assaying for arylsulphatase using the chromogenic sulphate X-SO$_4$ (Ohresser, Matagne et al. 1997). Similarly the *RSP3* gene (encoding flagellar radial spoke protein) can be used as a reporter when introduced into a paralysed flagella mutant background, *pf14*. In this case restored cell motility is used as the assay of promoter activity (Haring and Beck 1997). Protein localisation studies are possible by epitope tagging of cloned genes using a DNA sequence encoding the haemagluttinin (HA) epitope (Kozminski, Diener et al. 1993). Similarly, Jarvik et al. have designed a DNA cassette that allows the insertion of a 'mini-exon' into an intron of a cloned gene (Jarvik, Adler et al. 1996). This exon encodes an 18 amino acid epitope sequence which again allows the localisation of the tagged protein. However, the uses of reporters for promoter activity are limited, and the epitope tagging approach cannot be used to monitor gene expression *in vivo*.

The use of reporter molecules such as β-glucoronidase, luciferase or green fluorescent protein for gene expression and protein localisation studies in *Chlamydomonas* has been complicated by problems of heterologous gene expression in the organism. So far there have been no reports of successful expression of green fluorescent protein or firefly luciferase although some success has been attained with the bacterial *uidA* gene, encoding β-glucoronidase (GUS). The gene product of *uidA*, β-glucoronidase, can cleave a wide variety of artificial chromogenic substrates, providing a colourimetric assay for gene expression and the GUS gene fusion system has been used extensively as a reporter gene in plant and agricultural molecular biology (Jefferson 1989). Chloroplast transformants of *Chlamydomonas* containing transcriptional fusions between a *Chlamydomonas* chloroplast promoter region, *uidA* and the *3'* untranslated region of the chloroplast *petD* gene have been found to accumulate a high level of GUS activity (Sakamoto, Kindle et al. 1993). However,

so far attempts to express *uidA* in the nuclear genome of *Chlamydomonas* have not been successful. A desirable goal for molecular biological techniques in *Chlamydomonas* is the development of a suitable reporter molecule for gene expression in living cells. GFP has been used as a reporter in a variety of heterologous systems, and this chapter investigates the utility of GFP as a reporter for *Chlamydomonas*. The fluorescence of GFP results from cyclisation and oxidation of a Ser-Tyr-Gly sequence at amino acid residues 65-67 (Cody, Prasher et al. 1993). Detection of GFP in living cells thus only requires excitation by light at 395 nm or 470 nm. The fluorescence of GFP is stable and can be monitored *in vivo* in the absence of other proteins, substrates or co-factors. In contrast, the assay of GUS expression is cytotoxic (Jefferson 1989) and firefly luciferase requires the substrate luciferin (Ow, Wood et al. 1986; Millar, Carre et al. 1995). Another advantage of GFP is that it is relatively small (29.9 kDa) and can tolerate both N- and C- terminal fusions, lending itself to studies of protein localisation and intracellular protein trafficking (Kaether and Gerdes 1995). In addition, GFP mutants with shifted wavelengths of absorption and emission have been isolated (Delagrave, Hawtin et al. 1995; Heim, Cubitt et al. 1995) which permit simultaneous use and detection of multiple reporter genes (Yang, Kain et al. 1996).

Recent work has led to the successful expression of several bacterial genes in *Chlamydomonas*. This has demonstrated that the efficient expression of foreign genes in *Chlamydomonas* is feasible, and has provided insights into the different factors that mediate expression of transgenes in the alga (Lumbreras, Stevens et al. 1998). This chapter details attempts to express green fluorescent protein in the alga. Several vectors utilising various combinations of regulatory elements and GFP coding sequence were constructed. These constructs were transformed into *Chlamydomonas* and analysed for expression of GFP by fluorescence analysis, Southern analysis, western analysis and northern analysis.

4.2 Results

4.2.1 Construction of pnGFP

The GFP wild-type *gfp* sequence was amplified by PCR from a yeast expression vector and cloned into the *Chlamydomonas* nuclear expression vector, pSP105. pSP105 was developed in the Purton laboratory and comprises a pBS+ based plasmid containing the upstream and downstream sequences of the highly expressed *Chlamydomonas* nuclear gene *RBCS2*, together with a synthetic multiple cloning site. The upstream region of *RBCS2* comprises 740 bp of the *RBCS2* promoter and the downstream region 231 bp of the *RBCS2* putative polyadenylation signal, providing a suitable promoter, start codon and downstream processing for expression of a transgene gene in the nucleus of *Chlamydomonas* (figure 4.1). Amplification of wild-type *gfp* was achieved through design of primers to the 5' end of the GFP coding sequence incorporating a *Sal*I restriction site (GFP-N5') and to the 3' end of the coding sequence incorporating a *Bam*HI restriction site
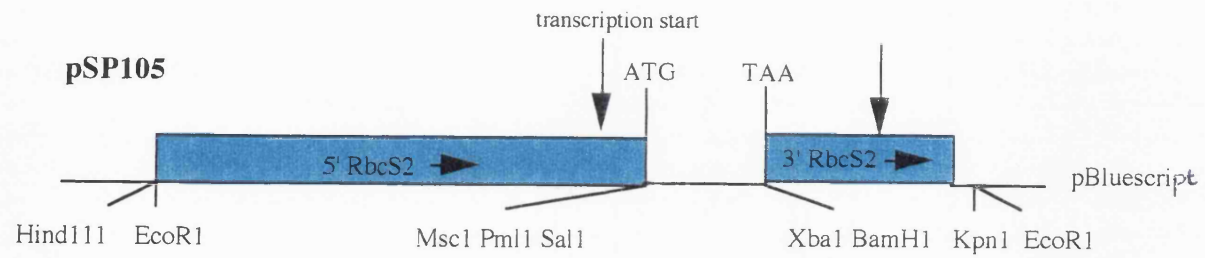
Figure 4.1   *Features of the nuclear expression vector pSP105*

(GFP-N3') in order to introduce suitable restriction sites for cloning into pSP105. For the sequences of the oligos, see Table 4 a.

The PCR product was blunt-cloned into the *Eco*RV site of pBluescript (Stratagene). The product was then excised from pBluescript using *Sal*I and *Bam*HI prior to cloning into the multiple cloning site of pSP105 cut with *Sal*I and *Bam*HI to create plasmid pnGFP (figure 4.6). Sequencing was carried out using primers raised to the promoter region of pnGFP in order to check that the *gfp* sequence was in frame with the translation start (figure 4.2).

4.2.2 Transformation of *Chlamydomonas* with pnGFP

The pnGFP plasmid was cut with *Pvu*II in order to linearise the transforming DNA. Linearisation of transforming DNA has been shown to aid integration of transforming DNA into the genome of *Chlamydomonas* and to minimise the occurrence of DNA rearrangements (Gumpel, Rochaix et al. 1994). The construct was then transformed into the nuclear genome of *Chlamydomonas* using the glass bead transformation method (Kindle 1990). The recipient strain for transformation was the cell-wall deficient, arginine-requiring strain *arg7.8* (mt⁻), CC-363. pnGFP was co-transformed along with the plasmid pARG7.8 which carries the gene *ARG7* encoding argininosuccinate lyase, (Purton and Rochaix 1995). Selection of transformants was by restoration of arginine-independent growth.

4.2.3 Fluorescence analysis of pnGFP transformants

GFP transformants in liquid culture were assayed for expression of GFP by measuring their emission at 510 nm compared to wild-type cells using a Perkin Elmer spectrophotometer with excitation at 395 nm. GFP transformants on solid media were assayed for fluorescence using the fluorescence video imaging system (see chapter 2). A 480 nm short-pass edge filter was used for excitation, and for emission a 510 nm narrow band pass filter (Ealing Electro Optics) was used to filter chlorophyll auto-fluorescence. No GFP fluorescence could be detected in the transformants. Detection of fluorescence in yeast expressing wild-type GFP (donated for use as a positive control by Dr Karen May of the Biology Department, UCL) was possible using both the Perkin Elmer spectrophotometer (figure 4.3, figure 4.4) and the fluorescence video imaging system.

4.2.4 Southern analysis of pnGFP transformants

Southern analysis of the pnGFP transformants was undertaken in order to verify that the GFP transformants harbour an intact copy of the wild-type *gfp* (figure 4.5). Four positive transformants were selected for a repeat of the fluorescence analysis, however once more fluorescence could not be detected in the transformants.
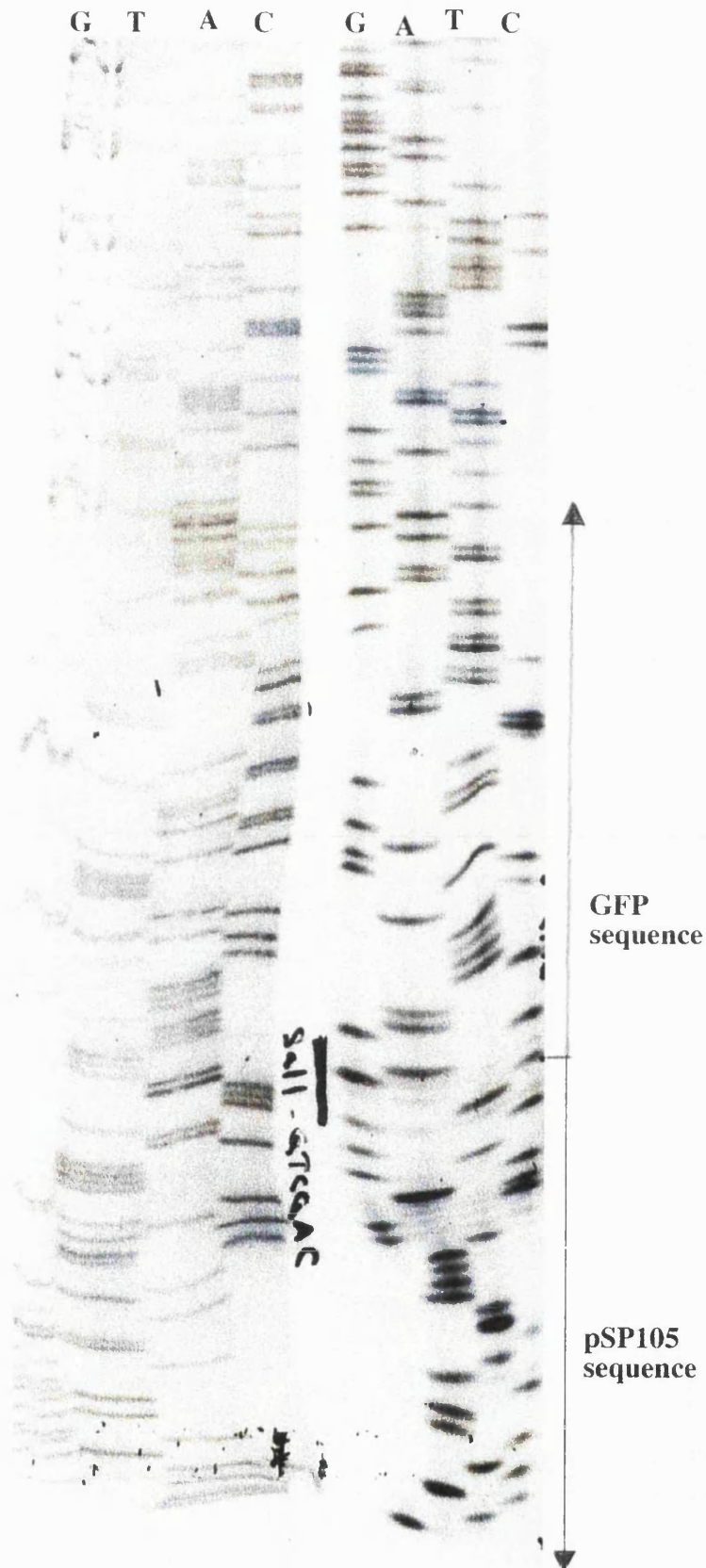
Figure 4.2   *Sequence across the promoter region of pnGFP*
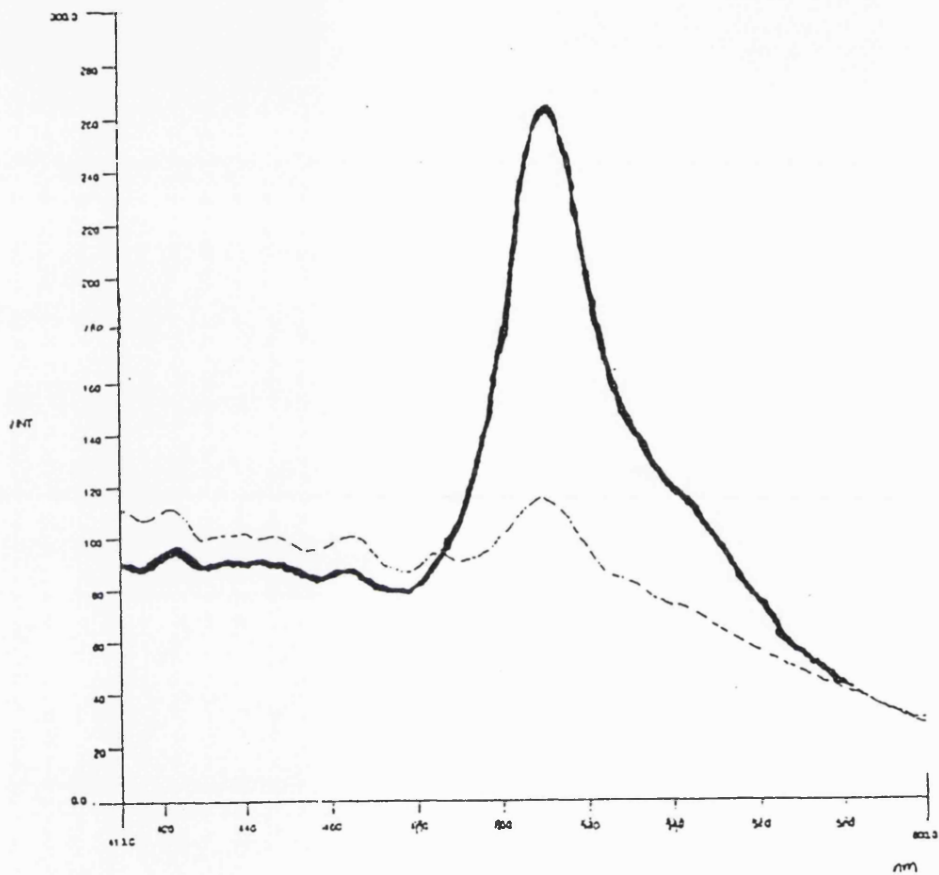
Figure 4.3 *Perkin-Elmer fluorescence emission spectra of* S. pombe *transformed with GFP. The solid line represents the fluorescence when expression is induced, the dashed line represents fluorescence when expression is repressed.*

99

a)
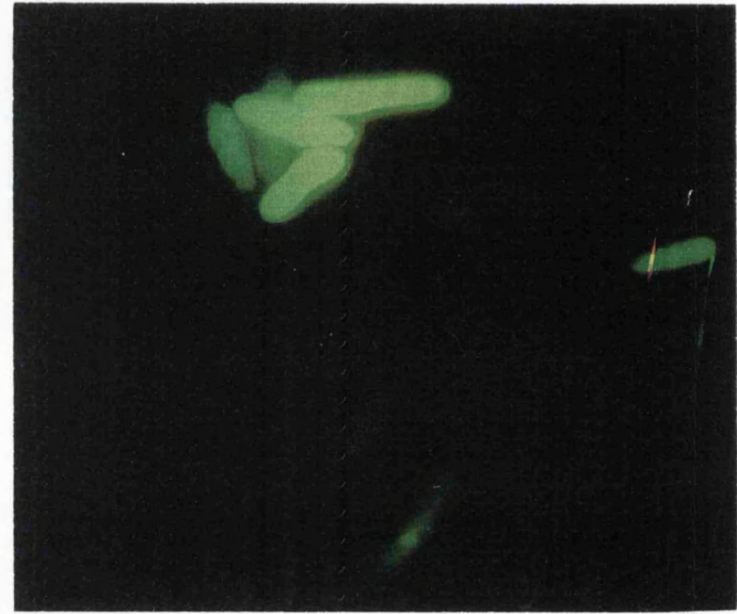
b)

Figure 4.4 *Phase contrast of a) wild-type* S. pombe *and b)* S. pombe *expressing GFP*
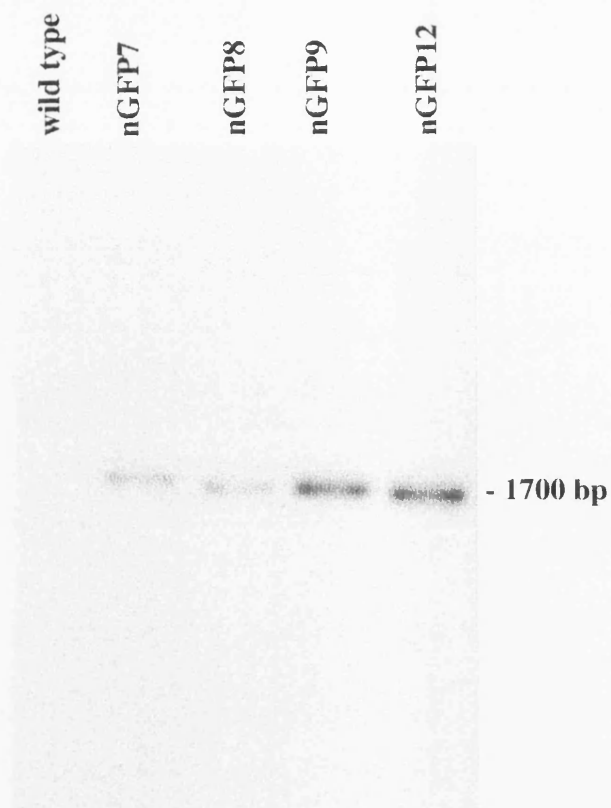
Figure 4.5 *Southern analysis of pnGFP transformants. 10 µg of total genomic DNA extracted from the pnGFP transformants and from wild-type cells were digested with EcoR1 and separated on a 1% agarose gel by electrophoresis. The DNA was transferred to a nylon filter by Southern blotting and probed with ^{32}P labelled GFP DNA. The filtered was washed for 2 x 20 minutes in 0.1% SDS, 2x SSC at room temperature, then for 2x 20 minutes in 0.1% SDS, 0.2 x SSC at 65 °C. The filter was then autoradiographed for 24 hours at -70 °C.*
A positive signal was obtained for all four of the pnGFP transformants. No GFP DNA could be detected for the wild-type negative control.

101

4.2.5 Truncation of the promoter of pnGFP

One reason proposed for lack of detection of GFP is that the protein is expressed, but at a level that is too low to be detectable by fluorescence imaging. Work in the Purton laboratory on the molecular mechanisms underlying the expression of foreign genes in *Chlamydomonas* has shown that the transformation efficiency of the *BLE* marker is improved several fold when the *BLE* coding region is fused to shorter regions of the *RBCS2* promoter than for the original *BLE* expression construct pSP108 (Lumbreras, Stevens et al. 1998). pSP108 utilises the same upstream and downstream regions of *RBCS2* as the vector pSP105 (figure 4.1). Various deletion derivatives of pSP108 were assayed for transformation efficiency in *Chlamydomonas*. The construct, pSP109, in which the promoter had been reduced to 180 bp of the original 740 bp *RBCS2* promoter sequence showed a three fold increase in efficiency of the *BLE* marker relative to the original pSP108 plasmid. The increase in transformation efficiency in the absence of this region may reflect an improved stability or rate of integration of the transforming DNA, or that a negative regulatory element is present in the regions of *RBCS2* upstream of the -180 site. The latter been reported independently by Cerruti et al. (Cerutti, Johnson et al. 1997). In theory it should be possible to achieve similar success with the expression of other transgenes.

The promoter of pnGFP was truncated from -740 to -180 bp as in construct pSP109. This was achieved through excision of a 560 bp *SmaI/HindIII* fragment from the pnGFP plasmid, followed by religation to create ptnGFP (figure 4.6). The construct was transformed into *Chlamydomonas* using the glass bead transformation method (Kindle 1990). The recipient strain for transformation was the cell-wall deficient, arginine requiring strain arg7.8 (mt⁻), CC-363. ptnGFP was co-transformed along with the plasmid pARG7.8 (Purton and Rochaix 1995) and selection of transformants was by restoration of arginine-independent growth. The resulting transformants were assayed for fluorescence. Once again, GFP fluorescence could not be detected in the transformants.

4.2.6 GFP variants with enhanced emission intensities

Wild-type GFP has several undesirable properties including low fluorescence intensity when excited by blue-light at the minor 470 nm peak, a significant lag in the development of fluorescence after protein synthesis, complex photoisomerisation and poor expression in certain plant and mammalian species (Cubitt, Heim et al. 1995; Chiu, Niwa et al. 1996). Mutants of GFP have been reported in which the excitation maximum has been shifted from 395 nm to around 490 nm. These proteins have been found to fluoresce with greater intensity than wild-type GFP when excited at 488 nm (Heim, Prasher et al. 1994; Delagrave, Hawtin et al. 1995). A mutant of GFP, S65T, in which the serine 65 of the chromophore has been changed to a threonine has been found to fluoresce six times brighter than the wild-type recombinant GFP when excited by blue light, and also to form a

functional chromophore more rapidly (Heim, Cubitt et al. 1995; Chiu, Niwa et al. 1996). A second mutant has been described which has a mutation of F64⇒L in addition to the

S65⇒T mutation, and has been found to exhibit 35 times brighter fluorescence relative to wild-type GFP in studies using equal amounts of protein (Cormack, Valdivia et al. 1996). These changes in the GFP coding sequence provide an enhanced GFP that increases the sensitivity of the reporter protein. It was decided to create a GFP plasmid for expression in *Chlamydomonas* that carries the S65T, F64L mutant version of GFP and to assay the transformed cells for expression of the enhanced form of GFP by fluorescence analysis.

4.2.7 Construction of pnGFP/S65T

A plasmid was constructed utilising the F64L, S65T codon changes to increase the fluorescence intensity of the recombinant GFP for expression in *Chlamydomonas*. This was achieved through re-engineering part of the GFP coding sequence encoding phenylalanine 64 and serine 65 using PCR directed mutagenesis. A primer incorporating the two codon changes was raised to the wild-type GFP coding sequence. The primer required a substitution of T for C in the first position of codon 64 for a change of F64⇒L, and a

substitution of T for A at the first position of codon 65 for a change of S65⇒T. An *Nco*I site was included at the beginning of the primer to facilitate cloning into the *Nco*I site present in wild-type GFP coding sequence. A second primer was raised to the 3' UTR of *RBCS2* downstream of the *Nco*I site. For the full sequences of the primers, see Table 4 a. PCR was carried out using 1 ng of pnGFP as template. The product was blunt-cloned into pZerO (InVitrogen) cut with *Eco*RV. The product was excised from pZerO using *Nco*I, prior to cloning into pnGFP to create the plasmid pnGFP/S65T (figure 4.6).

4.2.8 Fluorescence analysis of pnGFP/S65T transformants

The pnGFP/S65T construct was transformed into *Chlamydomonas* in the same way as the pnGFP and the ptnGFP plasmids, using restoration of arginine-independent growth as the selectable marker. The resulting transformants were subject to fluorescence analysis. As the S65T, F64L variant of GFP has a shifted excitation maxima, excitation was carried out at 488 nm. Emission was measured at 510 nm. Once again, GFP fluorescence could not be detected in the transformants.

4.2.9 Poor translation efficiency of *gfp* may be due to codon bias

It was decided that the most probable reasons for the lack of GFP fluorescence are that *gfp* is not expressed in *Chlamydomonas*, or that expression was at a low level such that
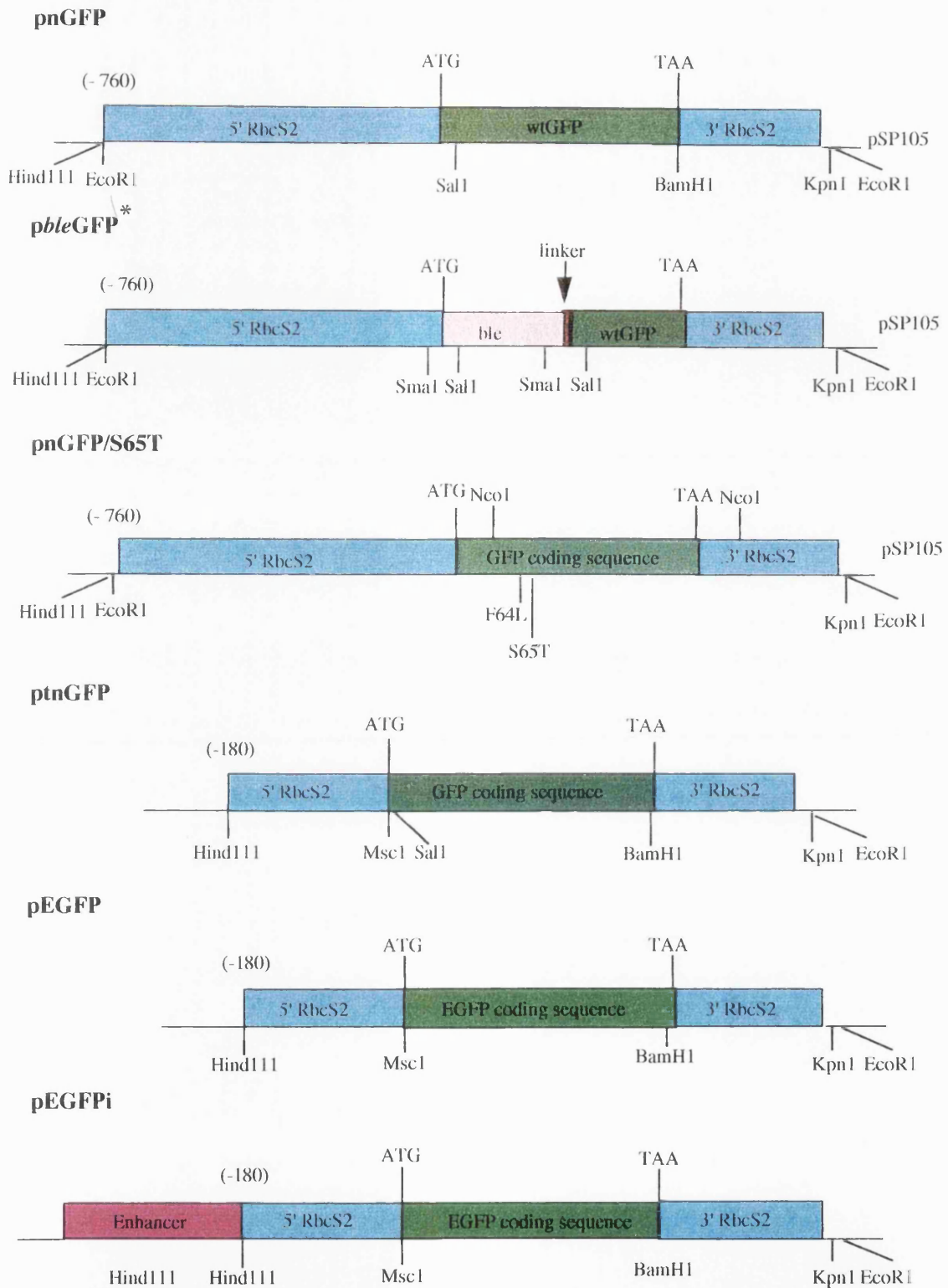
Figure 4.6   *Summary of various GFP expression constructs*

104

Name	Sequence, written 5'-3' (restriction sites underlined, degenerate base changes indicated in bold lettering)	Restriction sites	Codon changes (if any)
GFP-N5'	g atg agt aaa gtc gac gaa ctt ttc	SalI site	-
GFP-N3'	t tgg aag tct gga tcc tta ttt gta tag	BamHI	-
5'ble-GFP	g gtc gac gcc aag ctg acc agc gcc gtg ccg gtg	SalI	-
3'ble-GFP	g gtc gac ctt gga gcc gcc gtc ctg ctc ctc gcc cac	SalI	-
S65T	tt cca tgg cca aca ctt gtc act act ctc act tat ggt g	NcoI	F64⇒L, S65⇒T
3'RbcS2 rev	gta ccc gtc tca aat acg cc	-	-
EGFP 5'	ca cca tgg cca gca agg gcg agg ag	MscI	-
EGFP 3'	gt cgc gga tcc ttt act tgt aca gct cg	BamHI	-
EGFP int 5'	gac ggc aac tac aag acc cg	-	-
EGFP int 3'	gtt gta ctc cag ctt gtg ccc	-	-
GFP-C5'	aag ata aca atc atg agt aaa gga g	BspHI	-
GFP-C3'	aat tgg aag tct gca gat tta ttt gta t	PstI	-

Table 4 a   *Sequences of oligos used in the construction of GFP plasmids*

fluorescence is not detectable. A reason for low expression of GFP in heterologous systems is thought to be a lack of suitable controlling elements such as promoter, translation start site etc., or poor translation efficiency of mRNA in an environment characterised by a set of tRNAs that are different to those used in the jellyfish. It is known that the choice of codons in prokaryotic and eukaryotic genes is strongly biased, and there exist clear differences in codon usage between taxonomically distant organisms. These differences in codon choice do not affect the nature of the protein synthesised but may effect the level of expression (Ikemura 1981; Bennetzen and Hall 1982; Ikemura 1982). Difficulty in translating *gfp* in eukaryotic systems may be associated with the high A+U bias in the corresponding mRNA. Clusters of minor codons seem to be capable of modifying gene expression, especially when they are placed near the ATG initiation codon of the gene (Robinson, Lilley et al. 1984; Chen and Inouye 1990; Ueda, Taguchi et al. 1993). Earlier work with *Chlamydomonas* has shown that the alga exhibits a strong G+C codon bias (Rochaix 1995) but is capable of expression of the bacterial gene *BLE*, encoding a phleomycin resistance protein, to produce phleomycin resistant cells (Stevens, Rochaix et al. 1996). The *BLE* gene was chosen for this work due to its codon usage which closely matches that found in *Chlamydomonas* genes. It was decided to create a nuclear GFP expression vector incorporating a translational fusion between the *BLE* and the *gfp* sequence in order to decrease the occurrence of minor GFP codons near the translation start site and produce phleomycin resistant cells that also express GFP.

4.2.10 Construction of GFP plasmid, p*ble*GFP

A second GFP plasmid was constructed utilising a translational fusion between the bacterial phleomycin resistance gene, *BLE,* and *gfp*. Primers were raised to an expression plasmid encoding *BLE,* pSP108 (Stevens, Rochaix et al. 1996) such that the *BLE* coding region might be amplified by PCR. The primers were designed such that the PCR product has a *Sal*I restriction site at each end and encodes part of the GFP coding sequence at the 3' end (see Table 4 a). PCR was carried out using pSP108 as template. The PCR product was blunt-cloned into the *Eco*RV site of pZerO (Invitrogen), excised using *Sal*I, and cloned into pnGFP cut with *Sal*I to form the plasmid p*ble*GFP which has a translational fusion of *BLE* and GFP connected by a single codon (figure 4.6). In order to check that the *BLE* fragment had been cloned in the correct orientation, the positive minipreps were subject to restriction analysis using *Sma*I.

4.2.11 Transformation of *Chlamydomonas* with p*ble*GFP

p*ble*GFP was transformed into the nuclear genome of *Chlamydomonas* using glass bead transformation (Kindle 1990). The recipient strain for transformation was the cell wall

deficient *cw10* (mt-), CC-849. Selection of transformants was through resistance to phleomycin conferred by the *BLE* marker.

4.2.12 Fluorescence analysis of p*ble*GFP transformants

The positive transformants were subject to fluorescence analysis, but once more GFP fluorescence could not be detected. As the cells are resistant to phleomycin and therefore must express *BLE*, we would also expect to see expression of *gfp* provided the DNA is transcribed and correctly translated. The transformants were subject to western analysis to look for an increase in the size of the BLE protein relative to wild-type BLE due to the translation of the GFP coding sequence.

4.2.13 Western analysis of p*ble*GFP transformants

Western Analysis was carried out for the p*ble*-GFP transformants, using antibodies raised against BLE protein (Cayla, France). Expression of *BLE* was confirmed by western analysis in two of the positive transformants (figure 4.7). It was assumed that the remaining transformants had silenced *BLE* due to their maintenance on non selective medium. The native BLE protein is a small soluble protein of 13.5 kDa (Dumas, Bergdoll et al. 1994). For the positive control and two of the positive transformants a single band was seen at ~16 kDa, slightly larger than the expected size of the BLE protein, and a larger band was detected at ~28 kDa. These bands were thought to be due either to translation of part of the GFP coding sequence or to dimerisation of the BLE protein. The western was then probed with antibodies raised against GFP (Clontech) to look for GFP protein that had been denatured or had failed to form a fluorescent chromophore due to incorrect folding of the protein. No evidence of GFP expression was detected using the antibodies. It seems that translation of *BLE* is terminated by the *gfp* sequence, probably due to rare codon usage in the *gfp* sequence, and that the increased size of the BLE protein is due to dimerisation of the BLE protein.

4.2.14 N terminal codon changes can cause GFP fluorescence to fail

One of the most common reasons for non-fluorescence of GFP is failure of chromophore formation. In order to define further the primary structure requirements for chromophore formation and fluorescence in GFP, a series of N- and C- terminal GFP deletion variant expression vectors have been created using the polymerase chain reaction and scanned for fluorescence (Dopf and Horiagon 1996). Scanning of these deletion mutants revealed that amino acids 2-232 of a total of 238 amino acids in the native protein were required for the characteristic emission and adsorption spectra of native GFP, and GFP deletion variants lacking fluorescent activity showed no evidence of heterocyclic ring structure formation. These observations suggest that the primary structure requirements for the fluorescent
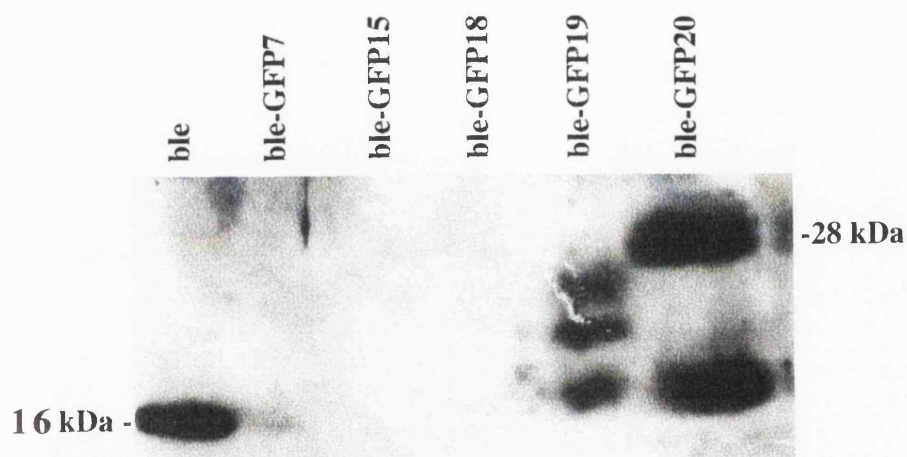
Figure 4.7 *Western analysis of pbleGFP transformants. Western analysis of pbleGFP transformants. Total protein extracts (measured on an equal chlorophyll basis equivalent to 10 µg of chlorophyll) from the positive pbleGFP transformants was loaded onto a tris-tricine gel. A* Chlamydomonas *strain transformed with pSP108 (encoding the BLE protein only) was used as a positive control. The proteins were separated by gel electrophoresis and transferred to a nitrocellulose filter. The resulting blot was probed with an antibody raised against BLE protein (obtained from Cayla, France). Primary antibody binding was detected using a horse radish peroxidase labelled secondary antibody and subsequent autoradiography.*

activity of GFP are relatively extensive and are compatible with the view that much of the primary structure serves an autocatalytic function. Among the deletion mutants characterised by Dopf and Horiagon was a mutant in which amino acids 2-8 had been removed from the coding sequence. The mutant failed to fluoresce. This evidence provides a possible explanation as to why GFP fluorescence had not been detected using the nuclear expression vector pnGFP and similarly the vectors pnGFP/S65T and p*ble*GFP, which are based on pnGFP. In the original *Chlamydomonas* expression vector, pnGFP, amino acid substitutions were made to the beginning of the amino acid coding sequence in order to accommodate the *Sal*I restriction site and facilitate cloning into the vector pSP105. In the light of evidence that amino acids 2 to 232 were required for fluorescence, it was revealed that the substitution of these amino acids is likely to effect the chromophore formation in the event of expression, and that the transformants would fail to fluoresce.

Original GFP sequence:

M   S   K   G   E   E   L   F   T   G   ....

pnGFP and pGFP/S65T expression constructs (amino acid substitutions shown in bold lettering):

M   S   K   **V**   **D**   E   L   F   T   G   ....

p*ble*GFP expression construct:

G   **S**   **K**   **V**   **D**   ....

No further analysis was undertaken with these transformants.

4.2.15 Construction of pEGFP

As previously mentioned, difficulty in translating the GFP gene in eukaryotic systems has been attributed to the high A+U bias in the corresponding mRNA. To improve the detection of GFP in transfected mammalian cells, a GFP variant has been constructed that is codon-optimised for higher expression in mammalian cells (Zolotukhin, Potter et al. 1996) and which contains the F64L,S65T chromophore mutations for enhanced fluorescence intensity (Cormack, Valdivia et al. 1996). Codon optimisation was achieved through construction of a synthetic GFP that substituted codons preferentially used in the human genome for the rarely used codons found in the original GFP gene. A total of 192 silent base substitutions were made through a strategy based on an overlap PCR extension method (Ho, Hunt et al.

1989), resulting in a G+C rich coding sequence. These changes in the GFP coding sequence provide an enhanced GFP (EGFP) that greatly increases the sensitivity of the reporter protein. In another experiment, strong expression in transgenic tomato and tobacco plants of the *Bacillus thuringenesis cryIA(b)* toxin, a gene with high A+U bias, was achieved through complete modification of the coding sequence. A synthetic gene with optimised codon usage was engineered and found to produce up to 100-fold higher protein levels than the wild-type gene (Perlak, Fuchs et al. 1991). Similarly, the re-engineered GFP gene sequence, with the favoured G+C rich codons of highly expressed human proteins, gives 20-fold higher GFP expression in maize leaf cells than the original jellyfish GFP sequence (Chiu, Niwa et al. 1996).

In the eukaryote *Chlamydomonas*, the nuclear DNA is found to be G+C rich (Rochaix 1995). The difference in codon choice and the relative abundance of tRNAs has been cited as a potential problem for the efficient translation of foreign mRNA lacking the codon bias found in *Chlamydomonas* genes (Kindle and Sodeinde 1994; Rochaix 1995). The EGFP codon usage was compared with the *Chlamydomonas* codon usage tables and found to compare favourably with that of *Chlamydomonas* nuclear genes. A nuclear expression vector was constructed utilising the EGFP coding sequence in the plasmid pSP109. Primers were raised to the EGFP sequence such that the coding sequence could be amplified from the pEGFP-N1 (Clontech) vector incorporating an *Msc*I site into the 5' EGFP primer and a *Bam*HI site into the 3' EGFP primer to facilitate cloning into pSP109 at the ATG start site. This requires a change of valine to alanine at amino acid 2 of the GFP coding sequence (T$\Rightarrow$C, G$\Rightarrow$C) for the 5' EGFP primer and changes of (G$\Rightarrow$T, G$\Rightarrow$A and C$\Rightarrow$G) for the 3' EGFP primer (see Table 4 a). PCR was carried out using pEGFP-N1 as template.

The PCR product was blunt-cloned into pZero cut with *Eco*RV, excised using *Msc*I and *Bam*HI, and cloned into pSP109 cut with *Msc*I and *Bam*HI to create the plasmid pEGFP (figure 4.6). The coding sequence was examined for errors by sequencing of the entire coding region. Sequencing was carried out using primers raised to the internal region of pEGFP in order to sequence the entire coding sequence (figure 4.8).

4.2.16 Inclusion of an enhancer element into GFP

Introns play an important role in the efficient expression of eukaryotic genes, and inclusion of an intron in a heterologous gene can significantly improve the expression of that gene. Such intron effects on transgenes have been described for other organisms including plants (Koziel, Carozzi et al. 1996) and the green alga *Volvox carteri* (Gruber, Kirzenger et al. 1996). This phenomenon has also been demonstrated in *Chlamydomonas*. A PCR fragment corresponding to the first intron of *RBCS2* was inserted into the *BLE* coding sequence at

5'RbcS2 -cgat gatgtttgat gggtatttga gcattgcaac ccttatccgg

aagccccttg gcccacaaag gtaggcxcca atgcaagcag ttcgcatgca

gcccxtggag cgttgccctc xttaaaccgg ccaggggcc tatgttcttt

acttttttac aagagaagtc gaactcaaca tcttaaatgg tgagcaaggg
                                          start
cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg

acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc

acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc

cgtgccctgg cccaccctcg tgaccaccct gacctacggc gtgcagtgct

tcagccgcta ccccgaccac atgaagcagc acgacttctt caagtccgcc

atgcccgaag gctacgtcca ggagcgcacc atcttcttca ggacgacgg

caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga
      5'EGFP int primer

accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg

gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc
      3'EGFP int primer

cgacaagcag aagaacggca tcaaggtgaa cttcaagatc cgccacaaca

tcgaggacgg cagcgtgcag ctcgccgacc actaccagca gaacacccccc

atcggcgacg gccccgtgct gctgcccgac aaccactacc tgagcaccca

gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc

tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac

taaaagagga tccgcgacta ctccgtgtta aatggaggcg ctcgttgatc
stop
tgagccttgc cccctgacga acggcggtgg atggaagata ctgctctcaa

gtgctgaagc ggtagcttag ctccccgttt cgtgctgatc agtcttttc

aacacgtaaa aagcggagga ttttgcaatt ttgttggttg tacgatctcc

gttgattttg gctxtttctc catgggcggg ctgggcgtat tgaagcggt-
3'RbcS2

Figure 4.8 *Sequence of pEGFP and pEGFPi. The EGFP coding sequence is shown in green and the upstream and downstream RbcS2 untranslated regions in blue.*

111

various sites (Lumbreras, Stevens et al. 1998). It was found that introduction of the *RBCS2* intron 1 into the middle of the *ble* coding region increased the transformation efficiency and that this effect can be further improved by placing the intron at the start of the coding region. The improvement is mediated in part by an enhancer element within the intron sequence. This element acts in an orientation-independent manner and is effective when placed upstream or downstream of the promoter, highlighting a possible role for introns as mediators of gene expression in the alga. A combination of the shortened *RBCS2* promoter and *RBCS2* intron 1 enhancer element results in a dramatic increase in transformation efficiency relative to the original pSP108 plasmid.

The construct pEGFP, which carries the truncated version of the *RBCS2* promoter, was further modified to include the 130 bp *RBCS2* enhancer element at the start of the *RBCS2* promoter sequence, creating the plasmid pEGFP*i* (fig 4.6). This was achieved through insertion of a blunt-ended PCR fragment of *RBCS2* intron 1 enhancer element into the *Hind*III site present at the -180 site of the promoter. The construct was checked for the presence of the enhancer element by restriction analysis.

A further factor thought to be associated with poor expression of transgenes in *Chlamydomonas* is the methylation of foreign DNA. In order to decrease the methylation state of the transforming DNA, the vector EGFP*i* was propagated in a strain of *E. coli*, JM110 lacking in the methylase enzymes *dam* and *dcm*.

4.2.17 Southern analysis of EGFP*i* transformants

Southern analysis was undertaken for cells transformed with pEGFP*i* in order to confirm that the transforming DNA had integrated into the nuclear genome of *Chlamydomonas* as a single copy and was stably maintained (figure 4.9). Only low levels of transformants harbouring the intact *EGFPi* coding sequence were recovered. It was thought that the reason for this could be that GFP is toxic to *Chlamydomonas*. As a test for toxicity of the EGFP transforming DNA, a comparison of numbers of transformants obtained by pARG7.8 and pEGFP*i* co-transformation and pARG7.8 transformation only, the results of which are presented in Table 4 b. It can be seen from the table that there is no significant difference in the numbers of transformants obtained by transformation with pARG7.8 only and by co-transformation with pEGFP and pARG7.8, therefore it was surmised that the EGFP coding sequence is not toxic to the cells and that lack of expression in the EGFP transformants is due to other factors.

4.2.18 Western analysis of pEGFP*i* transformants

Western analysis for those transformants found to contain an intact copy of GFP DNA was carried out using polyclonal antibodies against GFP (Clontech). Western analysis has an advantage over fluorescence analysis in that if the GFP protein has been expressed but has

	No.s of arg transformants per plate	No.s of arg and GFP co-transformants per plate
	199	153
	247	353
	310	194
	177	217
Average per plate	**233**	**224**

Table 4 b *Comparison of transformant numbers for pARG7.8 only and for pARG7.8 co-transformed with EGFPi*
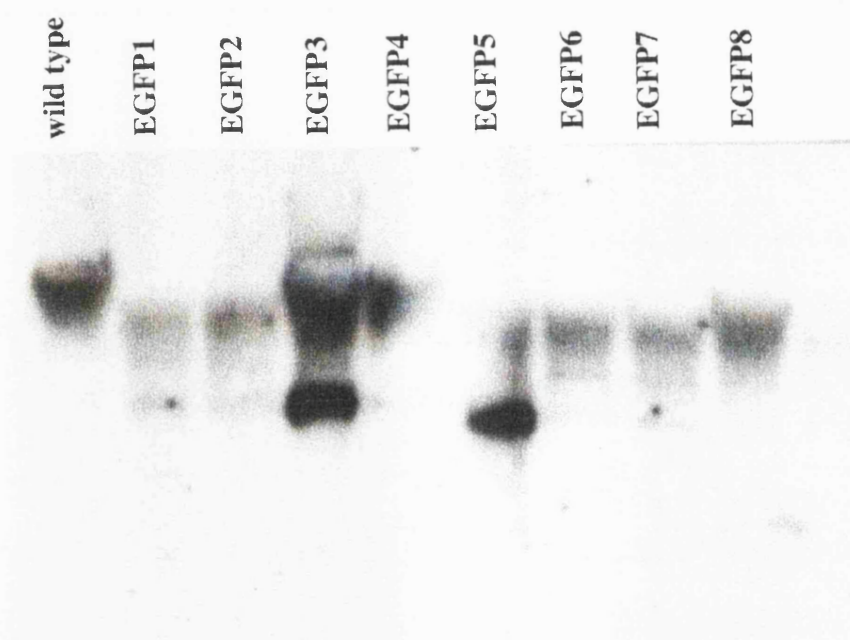
Figure 4.9 *Southern analysis of pEGFPi transformants. 10 μg of total genomic DNA extracted from the EGFPi transformants and from wild-type cells were digested with PvuII and separated on a 1% agarose gel by electrophoresis. The DNA was transferred to a nylon filter by Southern blotting and probed with ^{32}P labelled EGFP DNA. The filtered was washed for 2 x 20 minutes in 0.1% SDS, 2x SSC at room temperature, then for 2x 20 minutes in 0.1% SDS, 0.2 x SSC at 65 °C. The filter was then autoradiographed for 24 hours at -70 °C.*
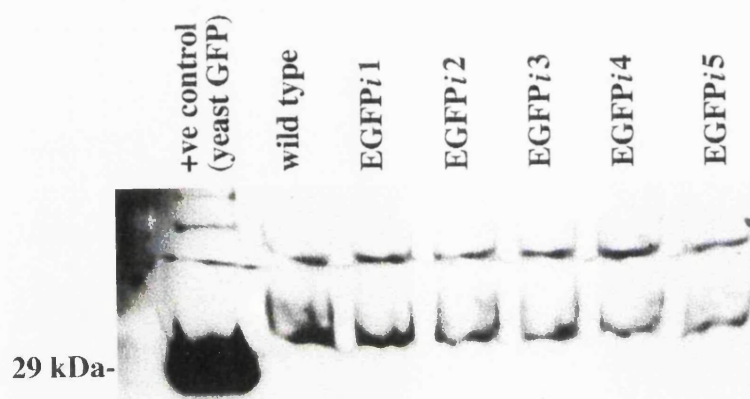
Figure 4.10 *Western analysis of EGFPi transformants. Total protein extracts (measured on an equal chlorophyll basis equivalent to 10 μg of chlorophyll) from the positive EGFPi transformants were loaded onto a tris-tricine gel. 5 μg of total yeast protein extract from a yeast strain expressing GFP was used as a positive control. The proteins were separated by gel electrophoresis and transferred to a nitrocellulose filter. The resulting blot was probed with an antibody raised against wild-type GFP (obtained from Clontech). Primary antibody binding was detected using a horse radish peroxidase labelled secondary antibody and subsequent autoradiography.*
Note that the EGFPi transformants show non-specific antibody binding.

not folded corectly or has been denatured, the protein should still be detectable. GFP extracted from yeast was used as a positive control. It was found that although the GFP antibodies gave a strong signal with the yeast GFP, no such signal could be detected in the *Chlamydomonas* transformants. Non specific binding to the *Chlamydomonas* protein extract gave a falsely positive result. Evidence for the same non specific binding was also seen in the negative control which comprises wild-type *Chlamydomonas* protein extract (figure 4.10).

4.2.19 Northern analysis of pEGFP*i* transformants

Northern analysis was conducted for the positive GFP transformants. No evidence for a GFP transcript could be detected.

4.2.20 Expression of GFP in the chloroplast of *Chlamydomonas*

Gfp expression in plants has been difficult to achieve and typically plants transformed with *gfp* have yielded very faint or no green fluoresence. As previously mentioned, the factors cited for the lack of expression of GFP in plants have been attributed to poor codon usage or incompatibility of controlling elements, as well as to the occurrence of a cryptic intron recognised by some plants leading to aberrant splicing of the GFP transcript (Haseloff, Siemering et al. 1997). Successful transient expression of GFP was initially attained in plant protoplasts (Hu and Cheng 1995; Niedz, Sussman et al. 1995; Sheen, Hwang et al. 1995). This has been attributed to the A+T bias of the chloroplast genome which is similar to the A+T bias of the wild-type GFP coding sequence. Targeting of GFP to intracellular organelles and removal of a cryptic intron have also been found to improve the fluorescence of GFP in plant cells (Haseloff, Siemering et al. 1997). Expression of GFP in plants has also been attained from virus based vectors (Casper and Holt 1996). In contrast to the nuclear genome of *Chlamydomonas*, which exhibits a high G+C bias, the chloroplast DNA of *Chlamydomonas* exhibits an A+T bias. Expression of a transgene in *Chlamydomonas* was initially achieved in the chloroplast genome of the organism, with the expression of the *aadA* cassette, conferring spectinomycin resistance. A GFP expression vector was constructed for the chloroplast in order to try to attain stable expression of GFP in the chloroplast DNA of *Chlamydomonas*.

4.2.21 Construction of pcGFP

The chloroplast expression plasmid, p72B, comprises a pUC8 based plasmid utilising upstream and downstream regions of the highly expressed chloroplast gene *rbcL* either side of the multiple cloning site (figure 4.11). These sequences provide an appropriate promoter, start codon and post-transcriptional processing signals for expression of a transgene in the chloroplast of *Chlamydomonas*.

116
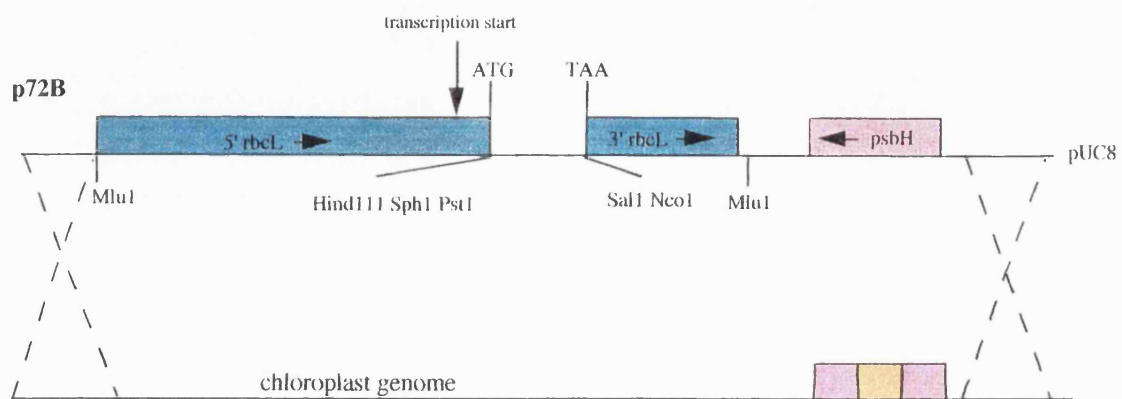
Figure 4.11 *Features of the chloroplast expression vector p72B and integration into the chloroplast genome of a* psbH *mutant in which the genomic copy of psbH is disrupted by* aadA *(shown in yellow) which confers spectinomycin resistance.*
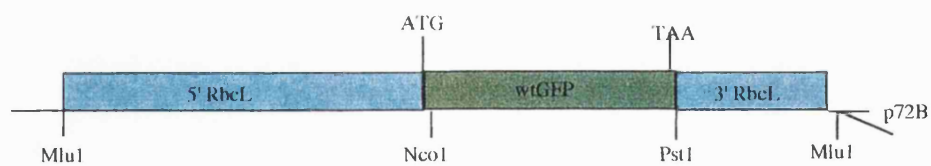
Figure 4.12 *pcGFP*
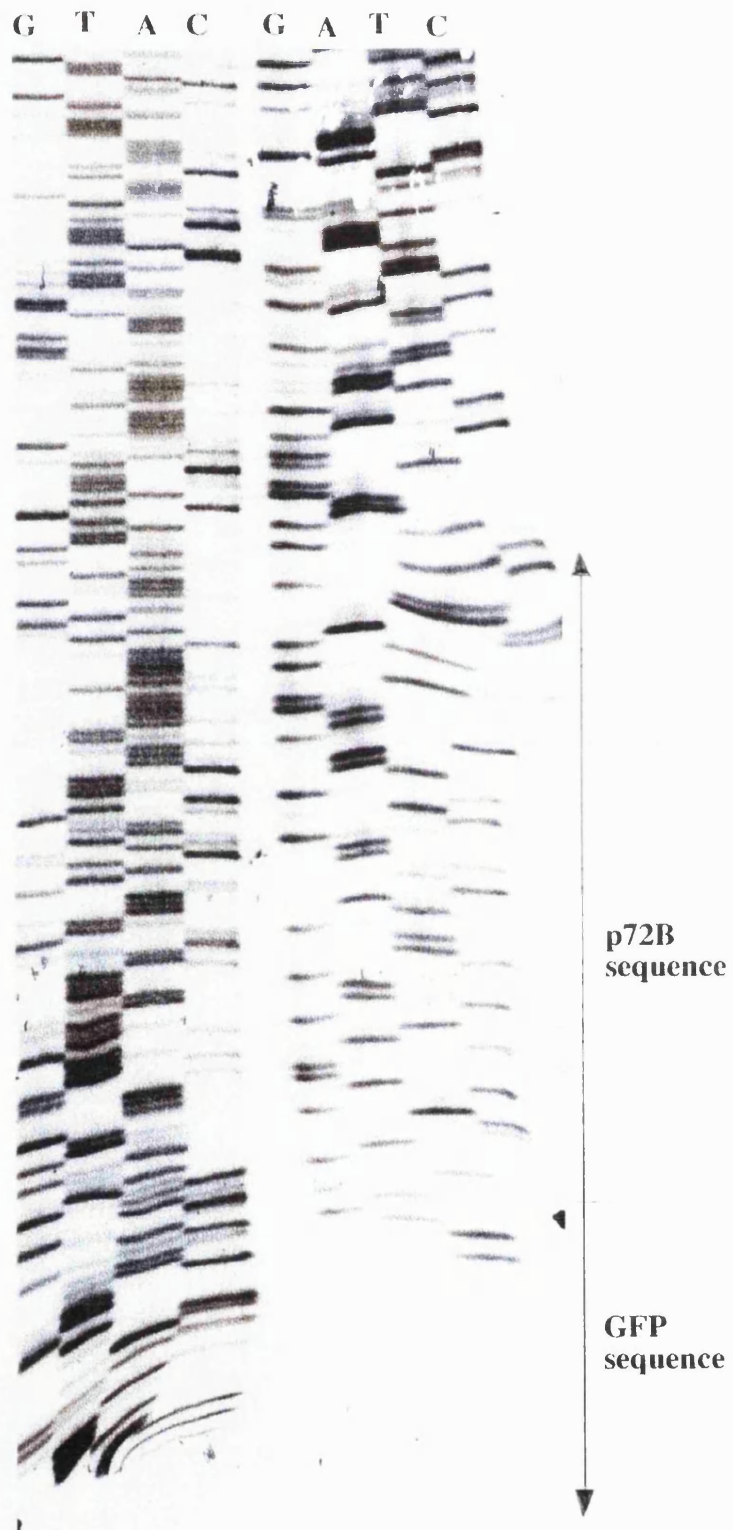
Figure 4.13   *Sequence across the promoter region of pcGFP*

118

The GFP coding sequence was amplified from the wild-type plasmid utilising primers raised to the 5' end of the GFP coding sequence incorporating a *Bsp*HI restriction site (GFP-C5'), and to the 3' end of the coding sequence incorporating a *Pst*I restriction site (GFP-C3'). The PCR product was blunt-cloned into the *Eco*RV site of pBluescript (Strategene). The product was then excised from pBluescript using *Bsp*HI and *Pst*I prior to cloning into the multiple cloning site of p72B cut with *Nco*I and *Pst*I to create plasmid pcGFP (figure 4.12). N.B. *Nco*I and *Bsp*HI have a compatible overhang with one another. Sequencing was carried out using primers raised to the promoter region of pcGFP in order to check that the *gfp* sequence was in frame with the translation start (figure 4.13).

4.2.22 Transformation of *Chlamydomonas* with pcGFP

The chloroplast expression construct pcGFP was transformed into the chloroplast genome by biolistic bombardment (Boynton, Gillham et al. 1989). The recipient strain for transformation is a mutant in which the endogenous *psbH* gene is disrupted by the *aadA* cassette, encoding spectromycin/spectinomycin resistance (O'Connor, Ruffle et al. 1998). The chloroplast expression vector p72B encodes a functional copy of the photosynthetic gene *psbH* _{down}stream of the *rbcL* promoter sequence to aid integration into the appropriate site of the chloroplast genome (figure 4.11). Selection of transformants was by restoration of full photosynthetic activity and sensitivity to spectinomycin (figure 4.14). In order to obtain homoplasmicity, the transformants were plated to single colonies on selective media for several generations.

4.2.23 Fluorescence analysis of pcGFP transformants

The positive transformants were subject to fluorescence analysis, but GFP fluorescence could not be detected.

4.2.24 Southern analysis of pcGFP transformants

Southern analysis was carried for the chloroplast transformants after several generations on selective media in order to confirm the presence of the GFP coding sequence and that the transformants had attained homoplasmicity (figure 4.15). It can be seen from the Southern blot that one of the cGFP transformants does not carry a copy of the GFP DNA.

4.2.25 Western analysis of pcGFP transformants

Western analysis for those transformants found to contain an intact copy of GFP DNA was carried out using polyclonal antibodies against GFP (Clontech). GFP purified from yeast was used as a positive control. It was found that although the GFP antibodies gave a signal with the yeast GFP, no such signal could be detected in the *Chlamydomonas* transformants. Non specific binding to the *Chlamydomonas* protein extract gave a falsely positive result.

Transformants

Untransformed strain
(PSII⁻, spec^R mutant)
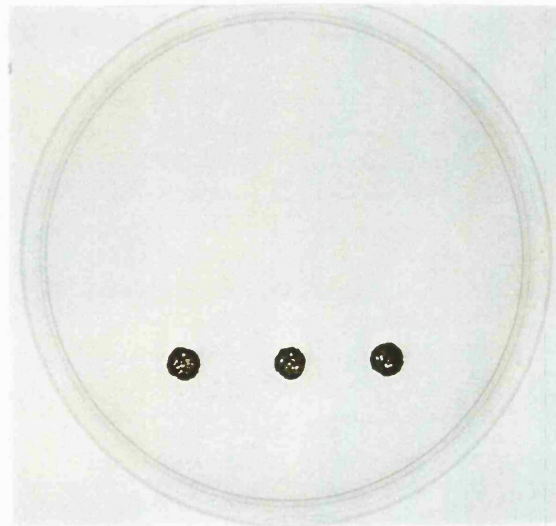
TAP + 100mM
Spectinomycin
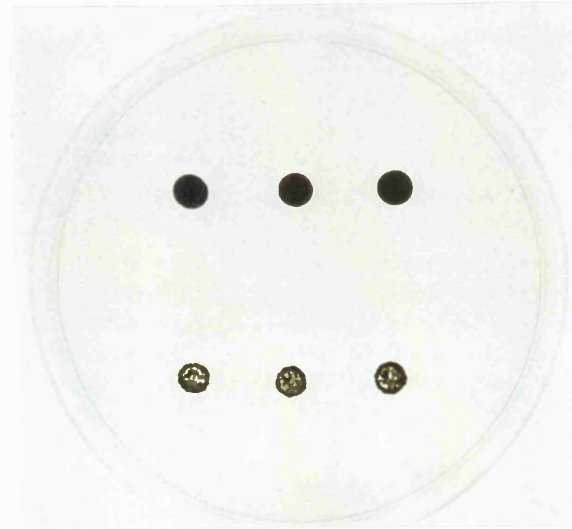
TAP only

Figure   4.14 *Chloroplast transformants showing restored sensitivity to spectinomycin*
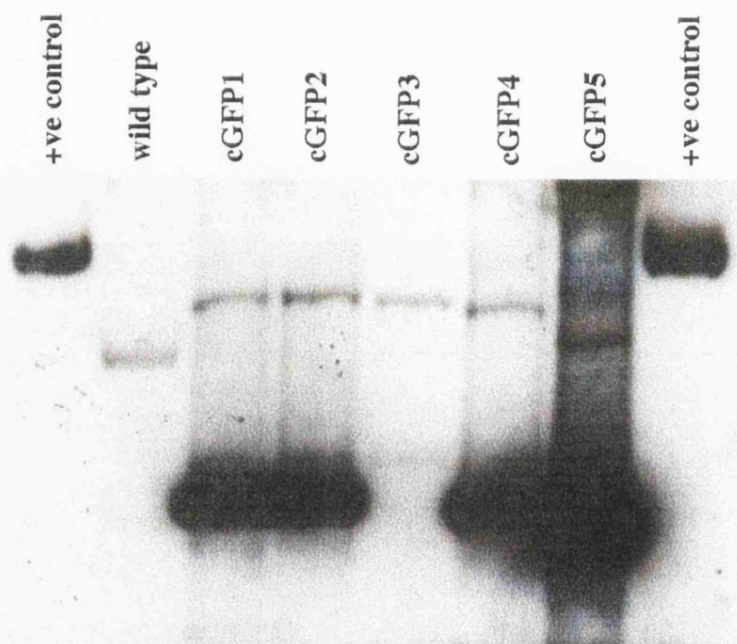
Figure 4.15 *Southern analysis of pcGFP transformants. 10 μg of total genomic DNA extracted from the EGFPi transformants and from wild-type cells were digested with PvuII and separated on a 1% agarose gel by electrophoresis. The DNA was transferred to a nylon filter by Southern blotting and probed with ^{32}P labelled GFP DNA. The filtered was washed for 2 x 20 minutes in 0.1% SDS, 2x SSC at room temperature, then for 2x 20 minutes in 0.1% SDS, 0.2 x SSC at 65 °C. The filter was then autoradiographed for 24 hours at -70 °C.*
No signal could be detected for the wild-type DNA, however a strong signal was detected for cGFP transformants 1, 2, 4 and 5, indicating that the transformants had attained homoplasmicity. No signal could be detected for cGFP3, indicating that the transformant had not integrated a copy of the GFP DNA. No further analysis was undertaken with this transformant.
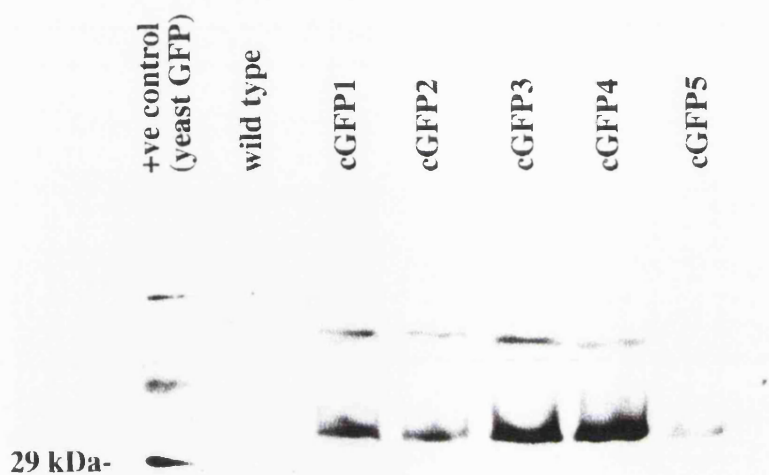
Figure 4.16 *Western analysis of pcGFP transformants. Total protein extracts (measured on an equal chlorophyll basis equivalent to 10 µg of chlorophyll) from the positive pcGFP transformants were loaded onto a tris-tricine gel. 5 µg of total yeast protein extract from a yeast strain expressing GFP was used as a positive control. The proteins were separated by gel electrophoresis and transferred to a nitrocellulose filter. The resulting blot was probed with an antibody raised against wild-type GFP (obtained from Clontech). Primary antibody binding was detected using a horse radish peroxidase labelled secondary antibody and subsequent autoradiography.*
Note that the pcGFP transformants show non-specific antibody binding. The same evidence for non-specific binding is also seen in the wild-type lane, although bleaching of the blot has occurred during photography.

Figure 4.17 *Northern analysis of positive pcGFP transformants, using ^{32}P labelled GFP cDNA. Total RNA from WT and cGFP transformants was separated on a 0.5% denaturing agarose gel, and transferred to nitrocellulose membrane. A single band can be seen at 3000 bp. It is not clear whether this is due to transcription run-on, as explained in the text, or to non-specific binding of the probe to ribosomal RNA.*

Evidence for the same non specific binding was also seen in the negative control which comprises wild-type *Chlamydomonas* protein extract (figure 4.16).

4.2.26 Northern analysis of pcGFP transformants

Northern analysis was conducted for the positive pcGFP transformants. Evidence for a transcript was detected for two of the transformants. However, the transcript was much larger than the expected size of the GFP transcript (figure 4.17). This was assumed to be due to failure of termination of the transcript during the transcription process. The positive transformants had been subject to western analysis, but no evidence for translation of the transcript into a mature GFP protein could be seen. Presumably, proper expression of GFP in the chloroplast of *Chlamydomonas* is curtailed by aberrant processing of the transcript.

4.3 Discussion

Although GFP has tremendous potential as a vital marker for continuously monitoring gene expression *in situ*, its utility in *Chlamydomonas* has still to be determined. Despite the many early successes of the use of GFP as a reporter in heterologous systems, expression of GFP has proved to be problematic in some systems. The main reasons put forward for non-expression of GFP are:

1. Incompatibility of the detection system, such that fluorescence cannot be detected for the positive control.

2. Formation of the GFP chromophore is temperature sensitive. In some cases *E. coli*, yeast and mammalian cells expressing GFP have shown stronger fluorescence when grown at lower temperatures. Another possible reason for lack of fluorescence is the requirement of the chromophore for molecular oxygen, therefore the cells must be grown in aerobic conditions.

3. GFP expression is too low to be detectable by fluorescence imaging.

4. The *gfp* DNA is transcribed and translated, but protein fails to become fluorescent or is denatured.

5. In some cases, it seems that GFP can be toxic to cells.

6. The gene is not expressed due to incompatibility of controlling elements, biased codon usage, silencing of introduced genes by methylation etc.

The failure to detect evidence of *gfp* expression in *Chlamydomonas* at any level is disappointing. A positive control in the form of GFP fluorescence in yeast was visible using the Perkin Elmer spectrophotometer and on plates using the fluorescence video imaging system. *Chlamydomonas* cells are grown in aerobic conditions at 25 °C, so temperature and oxygenation conditions should not affect formation of the chromophore. It is possible that GFP is expressed at a very low level, such that it is undetectable by fluorescence imaging, but western analysis confirmed this not to be the case. Polyclonal antibodies raised against GFP should detect the presence of the protein at a low level, even if the protein is partially denatured or fails to form the chromophore due to incorrect folding. A comparison of numbers of transformants produced by transformation of the cell wall deficient, arginine requiring strain *arg7-8 cw$_d$* (mt-), CC-363 with pARG7.8 only and co-transformation with pARG7.8 and pEGFP*i* yielded equal numbers. This would argue against GFP being toxic to the cells. From Southern analysis it can be seen that GFP DNA is integrated into the nuclear and chloroplast genomes respectively and stably maintained. However, northern analysis shows no evidence of a GFP mRNA transcript in nuclear transformants. It would seem, therefore, that control of GFP expression takes place at the transcriptional level. Either the GFP DNA is not being transcribed, or it is being transcribed but the resulting mRNA is rapidly degraded.

Heterologous expression of foreign genes in *Chlamydomonas* is inherently problematic, and there are only a few examples of successful expression of transgenes in the nuclear genome of *Chlamydomonas* so far (see Chapter 1). The factors thought to affect expression of transgenes in *Chlamydomonas* include:

1. Incompatibility of controlling elements.

2. Silencing of introduced gene by methylation.

3. Biased codon usage.

The truncated *RBCS2* promoter together with an enhancer element upstream of the transcription start point provides an efficient strong promoter in *Chlamydomonas* from which expression of both endogenous and transgenes can be driven (Cerutti, Johnson et al. 1997; Lumbreras, Stevens et al. 1998). Similarly, the chloroplast expression vector p72B provides a strong promoter for expression of a gene in the chloroplast, and there are examples of this promoter being used to drive expression of both endogenous and transgenes in the chloroplast of *Chlamydomonas* (Erickson 1996). The p*ble*GFP transformants show evidence of *BLE* expression both by selection on phleomycin and in western analysis, but no evidence of *gfp* expression. This is in line with the theory that the

promoter and downstream processing signals are suitable for driving expression of GFP. The reason that GFP expression cannot be detected is intrinsic to the GFP coding sequence. It is possible that the GFP transforming DNA is silenced by heavy methylation. DNA methylation is correlated with decreased expression of many eukaryotic genes, a process which has been extensively studied in animal systems (Holliday 1987). DNA methylation also appears to be involved in the regulation of plant gene expression. Plant genomes are highly methylated at both CG and CNG sites (Gruenbaum, Naveh-Many et al. 1981). In contrast actively transcribed regions are hypomethylated compared with the bulk DNA (Antequera and Bird 1988). The methylation state of the transforming DNA could be investigated by Southern analysis using restriction enzymes that allow discrimination between methylated or non-methylated DNA sequences (Renckens, Greve et al. 1992) or by treatment of cells with 5-azacytosine, a potent inhibitor of DNA methyltransferases which has been shown to result in demethylation and subsequent expression of T-DNA genes in plants (Klaas, John et al. 1989).

Control of gene expression can also result from regulation at later steps than transcription such as mRNA degradation (Sullivan and Green 1993), splicing (Luherson, Taha et al. 1994) or translation. The ability to recognise and rapidly degrade certain transcripts provides a means of minimising the production of wasteful and possible deleterious truncated protein. Messenger RNAs carrying early termination codons resulting from nonsense or frameshift mutations have been shown to be subject to nucleolytic degradation *in vivo*. Plants (Hoof and Green 1996), yeast (Leeds, Peltz et al. 1991), mammals (Maquat 1995), *C. elegans* (Pulak and Anderson 1993) and *Chlamydomonas* (Li, Quinn et al. 1996) all appear to have evolved mechanisms to accelerate the decay of certain transcripts containing early nonsense or frameshift mutations, thus minimising the production of truncated proteins and improving evolutionary fitness. Hoof and Green examined the effects of premature nonsense codons introduced into the coding sequence of the bean phytohemagglutinin gene (PHA). Measurement of mRNA decay rates in stably transformed cell lines demonstrated that premature nonsense codons markedly destabilised the mRNA, reflected by decreased accumulation of transcripts containing premature nonsense codons in higher plants (Hoof and Green 1996). In *Chlamydomonas*, mRNAs carrying early termination codons resulting from frameshift mutations have also been shown to be subject to nucleolytic degradation. A plastocyanin deficient mutant of *Chlamydomonas* was found to carry a frameshift mutation in the plastocyanin gene, *Pcy1*, resulting in loss of *Pcy1* mRNA relative to wild-type. The lack of *Pcy1* mRNA was attributed to increased degradation of rather than decreased synthesis since the mRNA can be stabilised by treatment of cells with

anisomycin, which are known to stabilise decay prone messages. Rare codons can also effect the rate of gene expression especially if they occur near the ATG initiation codon (Robinson, Lilley et al. 1984; Chen and Inouye 1990; Ueda, Taguchi et al. 1993). If the

transcriptional level of the coding sequence is presumed to be invariant, then this effect most likely occurs by control of the abundance of mRNA. Ueda et al. (1993) found that introducing a rare TTA^{LEU}, a codon rarely used by *Streptomyces* in the highly expressed subtilisin inhibitor (SSI) gene from *S. albogriseolus* resulted in a considerable reduction of SSI production by *S. lividans* until late in growth. This is in agreement with the transcriptional regulation of the *bldA* gene, which encodes for the rare TTA codon in *S. coelicolor* (Lewski, Mah et al. 1993).

The effect of a premature nonsense codon on mRNA stability can be greatly influenced by its position on the transcript. Termination of translation 80% through the coding region does not trigger rapid decay, wheras occurrence of a nonsense codon earlier in the transcript does (Hoof and Green 1996). The information that allows such an mRNA to be recognised as a candidate for degradation cannot be contained in the codon itself. There must be an additional *cis*-acting element contained in the transcript, or it may be simply the length of the translated or untranslated region that renders the mRNA unstable. Slow translation of a sequence due to the presence of rare mRNAs may also affect the rate on mRNA turnover. It could be that the transcription of *gfp* in *Chlamydomonas* takes place at a normal level but due to the presence of unfamiliar codons the resulting mRNA is recognised as foreign and subject to rapid degradation, hence the inability to detect the presence of a transcript in northern analysis.

Heterologous genes may contain processing and or instability signals that have a deleterious impact on gene expression. Another possible reason for the lack of a *gfp* transcript is that *gfp* mRNA processing is curtailed due the presence of a splice site in the transcript. In some plant systems it has been found that non-expression of *gfp* has been due to the existence of a cryptic intron (Pang, DeBoer et al. 1996; Haseloff, Siemering et al. 1997; Rouwendal, Mendes et al. 1997). For example, in the model plant *Arabidopsis* proper expression of *gfp* is curtailed due to aberrant mRNA processing. An 84 nucleotide cryptic intron is efficiently recognised and excised from transcripts of the GFP coding sequence. The cryptic intron contains sequences similar to those required for recognition of normal plant introns. The coding sequence was modified to mutate the intron and to restore proper expression in *Arabidopsis* (Haseloff, Siemering et al. 1997). All small G proteins (SGPs) in *Chlamydomonas* possess a very similar array of structural and functional domains, and they have been used for examining the relationships postulated to exist between the exon-intron structure of genes in *Chlamydomonas*. A consensus sequence consisting of a conserved GT-AG border to the intron and a pyrimidine-rich tail has been identified (Dietmaier and Fabry 1994). A search for possible internal splicing consensus motifs (e.g. lariat consensus sequence) using 116 different introns of *Chlamydomonas* for which sequences exist found no evidence for such motifs (Liss, Kirk et al. 1997). The same set of introns gave a pyrimidine content of 71 %, whereas the overall average pyrimidine content of

Chlamydomonas is 53 %. This is in agreement with the findings of Dietmaier and Fabry. If an internal splice site occurs within the EGFP coding sequence, especially early in the sequence then any mRNA produced may be too short to be recognised in northern analysis, or again rapidly degraded due to lack of post transcriptional factors thought to stabilise the mRNA, such as the poly-A tail.

Removal of minor codons or precise alteration of processing signals can be difficult due to the affects of the surrounding nucleotides and the imprecise nature of processing signals. Each minor change may contribute only a small improvement in expression. Rather than carry out multiple rounds of mutagenesis to change several scattered sites, it may be easier to construct a synthetic gene of the desired sequence, providing the opportunity to remove all potentially deleterious processing signals and minor codons (Koziel, Carozzi et al. 1996). Perhaps the answer to finding a way to express GFP in *Chlamydomonas* lies in the construction of such a synthetic gene incorporating favoured *Chlamydomonas* codon usage together with removal of any putative splice sites in the coding region. Such an approach has been attempted, but was unsuccessful [Peter Hegemann -personal communication]. It should be noted however that changes in DNA sequence involve changing the primary and secondary structure of the mRNA which in turn can effect mRNA stability (Lammertyn, Mellaert et al. 1996) and construction of such a synthetic gene may have an unpredictable effect on expression levels.

Chapter 5
GENE TARGETING AND ANTISENSE DOWN-REGULATION OF NUCLEAR GENES IN *CHLAMYDOMONAS REINHARDTII*

Chapter 5:

Gene targeting and antisense down-regulation of nuclear genes in *Chlamydomonas reinhardtii*

5.1 Introduction

5.1.1 The search for a negative selectable marker for *Chlamydomonas*

Gene targeting involves replacement of an endogenous gene with a modified or mutated version of that gene. It can be used to investigate the effects on phenotype of gene replacement with a non-functional copy ('gene-knockout') or used to produce site-directed modifications to a given gene. Gene targeting is straightforward in several model systems (e.g. yeast, *Dictyostelium*) because of the high frequency of homologous recombination that occurs in these organisms. However, in model systems such as mice, where homologous recombination occurs only at a low level, a negative selection stategy is required in order to obtain rare gene targeting events (Mansour, Thomas et al. 1988; Capecchi 1989). The strategy involves the use of a transformation construct in which a lethal gene is inserted into the vector along with the target sequence, which is usually a disrupted or mutated version of an endogenous gene. The lethal gene will have its own promoter and regulatory elements and if integrated into the genome, will be activated, causing cell death (figure 5.1). In the case of integration of the target DNA into its homologous site, crossover occurs between the two similar sequences i.e. only the target DNA will integrate and the remaining vector with its lethal gene will be degraded.

Such an approach could be adapted for use in *Chlamydomonas*, to provide a negative selectable marker for homologous recombination. The low frequency of homologous recombination in the nuclear genome of *Chlamydomonas* has prevented the development of routine methods for gene targeting in the alga (Sodeinde and Kindle 1993). At present, gene knockout mutants may be obtained using an indirect approach involving screening large numbers of transformants that display the expected mutant phenotype for those transformants in which the target gene has been disrupted as a result of a rare homologous recombination event (Wilkerson, King et al. 1995; Khrebtukova and Spreizer 1996; Myster, Knott et al. 1997). The only convincing example of a targeted gene knockout involved the disruption of a gene (*NIT8*) necessary for nitrate assimilation, and required a second selective regime to identify the few *CRY1-1* transformants affected in the assimilation pathway (Nelson and Lefebvre 1995).

In order to design a negative selectable marker for *Chlamydomonas* it is first necessary to find a lethal gene which will act as a 'suicide' marker for negative selection. One strategy for the isolation of such a gene is mutant rescue of cells resistant to the fluoro-analogue of a metabolic substrate. Substitution of certain metabolic substrates for their fluoro-analogues

130
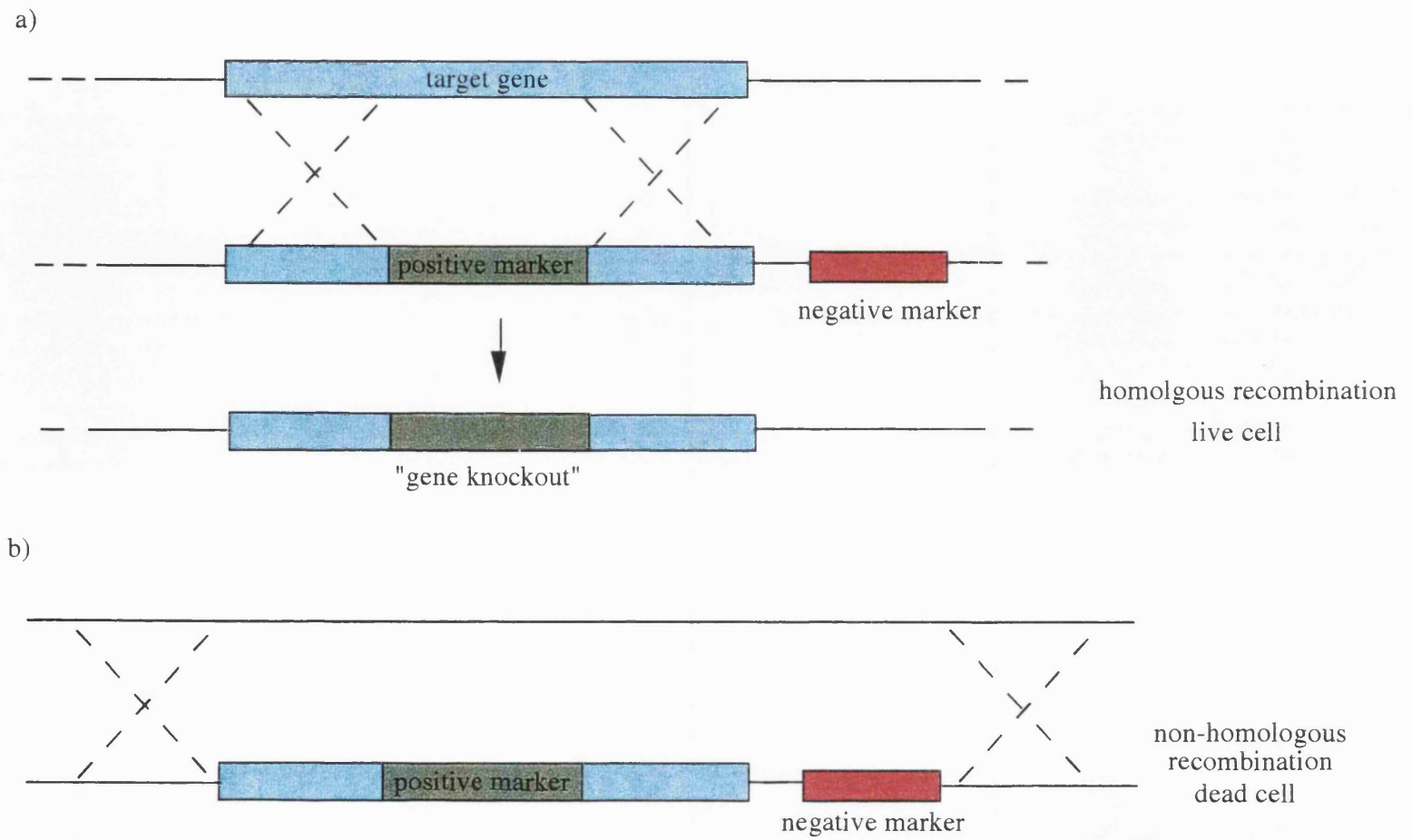
Figure 5.1   *A 'positive-negative' selection strategy*

Strategy:

$$X \xrightarrow[\text{enzyme}]{} Y$$

But:

$$\text{F-X} \xrightarrow[\text{enzyme}]{} \text{F-Y} \quad \text{Toxic!}$$

In resistant mutant:

$$\text{F-X} \xrightarrow[\text{enzyme}]{} \text{F-Y} \quad \text{No product}$$
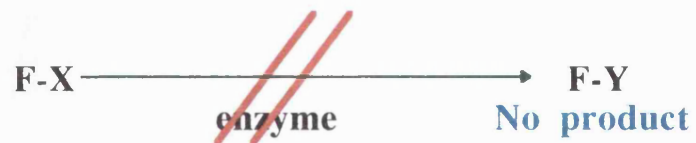
Rescue of mutants with wild-type enzyme:

$$\text{F-X} \xrightarrow[\text{enzyme}]{} \text{F-Y} \quad \text{Toxic !}$$
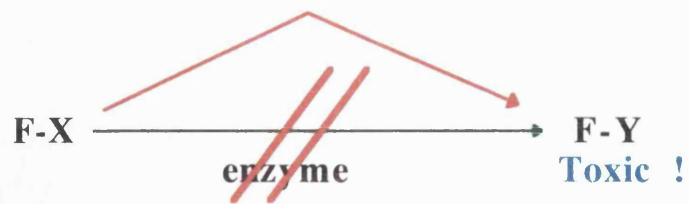
Figure 5.2 *Fluoro-analogue resistance selection strategy*

can cause cell death. The fluoro-analogue is assimilated into the metabolic pathways of the organism and its toxic product causes the cells to die. Mutants defective in key enzymes of the metabolic pathway will be unable to utilise the toxic fluoro-analogue, and these cells will survive. Fluoro-analogue resistance to isolate mutants affected in the control or function of a particular enzyme has been previously used in a range of organisms from *E. coli* to *Arabidopsis thaliana* (Hodson and Gresshoff 1987). Complementation of a resistant mutant with the cloned wild-type gene restores the wild-type phenotype and the cell once more becomes sensitive to the fluoro-analogue (figure 5.2).

The isolation of a fluoro-analogue resistant mutant in *Chlamydomonas*, followed by the cloning of the wild-type gene would provide a negative selection strategy for homologous recombination. The transformation vector would include the wild-type gene down stream of the target gene sequence. Cells that do not integrate the wild-type gene will be able to survive on the toxic substrate. Cells that integrate the target sequence but not the toxic marker gene should be enriched in transformants resulting from homologous recombination, although there will also be a background of non-homologous transformants in which the negative selectable marker has failed to integrate, has been subject to rearrangements or mutations during integration, or is poorly expressed in the transformant. This approach has the additional advantage that it avoids the added complications of heterologous gene expression in *Chlamydomonas*. Mutant rescue using cloned wild-type genes is an approach that has already been widely exploited in the alga for selection of transformants (Debuchy, Purton et al. 1989; Fernandez, Schnell et al. 1989; Diener, Curry et al. 1990; Kindle 1990; Mayfield and Kindle 1990).

Two pathways were investigated in which substitution of a substrate for its fluoro-analogue causes cell death in *Chlamydomonas;* that for UMP biosynthesis, and that for acetate utilisation.

i) The UMP biosynthesis pathway

The final enzyme in the synthesis of UMP (uridine monophosphate) is catalysed by the enzyme orotidine-5'-monophosphate decarboxylase (OMPD) which converts OMP to UMP (figure 5.3). This *de novo* pyrimidine biosynthetic pathway has been well studied genetically and enzymatically in several organisms, ranging from bacteria to mammals including yeast, *Dictyostelium*, Drosophila and *Arabidopsis* (Nasr, Betauche et al. 1994). The coding gene for OMPD has been cloned and characterised from a variety of organisms from *E. coli* to *Homo sapiens* (Benito, Diaz Minguez et al. 1992) (figure 5.4). The enzyme sequence appears to be well conserved during evolution, possessing stretches of highly conserved amino acids (Radford and Dix 1988).

5-fluoro-orotic acid (5-FOA) is a fluoro-analogue of orotic acid which is converted to the toxic intermediate 5-fluoro-UMP in wild-type cells (Boeke, Lacroute et al. 1984). Mutants

Figure 5.3 *Diagram of the* de novo *pyrimidine synthesis pathway. The key to the enzymes involved in uracil biosynthesis are as follows:*

1.3.1.2 - *dihydropyrimdase,* 1.17.4.1 - *ribonucleoside-diphosphate reductase,* 2.1.1.45 - *thymidylate synthase,* 2.4.2.9 -*uracil phosphorbosyltransferase,* 2.4.2.10 - *orotate phosphoribosyltransferase,* 2.7.4.4 - *nucleoside triphosphate,* 2.7.4.6 - *nucleoside-diphosphate kinase,* 3.5.1.6 - *β-ureidoproprionase,* 3.5.2.2 - *dihydropyrimidase,* 3.5.2.3 - *dihydro orotase,* 3.5.4.12 - *dCMP diaminase,* 3.6.1.15 - *nucleoside triphosphate,* 4.1.1.23 - *orotidine-5'-phosphate decarboxylase,* 6.3.4.2 - *CTP synthase*

Figure 5.4 *Comparison of UMP biosynthetic pathway in different organisms*

that lack OMPD activity are resistant to 5-fluoro-orotic acid but require uracil. Complementation of FOA resistant/uracil requiring mutants using the OMPD coding sequence provides positive selection that has been applied to the identification of uracil/uridine auxotrophs in various species (figure 5.5). Evidence for the existence of the UMP biosynthesis pathway in green algae includes the fact that *Chlamydomonas* mutants resistant to 5-FOA have been recovered [Scott Bingham-personal communication] and 5-FOA resistant, uracil requiring mutants have been isolated from diatoms, a large and extremely diverse group of microalgae found in sea, freshwater and terrestrial environments (Dunahay, Jarvis et al. 1996).

There are a number of ways in which an FOA resistant mutant of *Chlamydomonas* could arise:

1. Mutation of the gene encoding OMPD. Such mutants would be unable to assimilate 5-FOA and would require uracil.

2. Mutation of the gene encoding OMPD, where the organism is able to manufacture uracil by a previously uncharacterised pathway that does not involve synthesis from uridine monophosphate. Mutants will have a 5-FOA resistant phenotype, but will not require uracil.

3. Mutation in the uptake of 5-FOA. Such mutants would be resistant to 5-FOA as it can not be assimilated by the organism. Again they would not be uracil requiring.

A strategy for the development of a selectable marker for *Chlamydomonas* would be the isolation of 5-FOA resistant mutants in *Chlamydomonas* that are also uracil requiring due to lack of OMPD activity, followed by cloning of the wild-type OMPD gene. Complementation of the mutant with the wild-type gene would provide either positive or negative selection strategies - positive selection for uracil-independant transformants and negative selection for FOA sensitivity.

ii) The acetate utilisation pathway of *Chlamydomonas*

A second strategy for the development of a negative selectable marker involves the acetate utilisation pathway in *Chlamydomonas*. Acetamide ($CH_3.CO.NH_2$) can serve as both a carbon and nitrogen source for *Chlamydomonas*. It is hydrolysed by the enzyme acetamidase to ammonium (nitrogen source) and acetate (carbon source). This enzyme activity is induced by acetamide and urea, and repressed by ammonium. Acetamide also induces the urea metabolising pathway in which two enzyme activities (urea carboxylase and allophanate hydrolase) convert urea to ammonium and carbon dioxide (figure 5.6).

orotidine 5'
monophosphate $\xrightarrow[\text{ompd}]{}$ uridine 5'
monophosphate

Also:

uracil $\xrightarrow{}$ uridine 5'
monophosphate

But:

5-fluoro-orotic
acid $\xrightarrow[\text{ompd}]{}$ fluoro-uracil

In resistant mutant:

uracil

$\downarrow$

5-fluoro-orotic
acid $\xrightarrow[\text{ompd}]{\;\;//\;\;}$ UMP
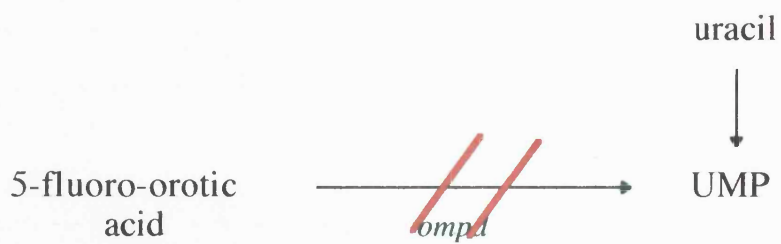
Figure 5.5 *The 5-FOA resistance strategy*

Fluoro-acetamide (FAM) kills wild-type cells under appropriate conditions by a suicide mechanism in which it is converted to fluoro-acetate and thence to fluoro-citrate, an inhibitor of the TCA cycle. Toxicity is dependant on: i) the presence of urea in the medium (to induce acetamidase activity); ii) the absence of ammonium in the medium (to prevent repression of acetamidase activity) and the absence of acetate in the medium (to facilitate the uptake/conversion of fluoro-acetate).

There are a number of ways that resistance to FAM could arise:

1. Mutation of the acetamidase gene. Such mutants would not be able to use acetamide as a carbon or nitrogen source, but otherwise would have a wild-type phenotype.

2. Mutation in a regulatory gene involved in the induced expression of the acetamidase gene. As above, such mutants would be unable to use acetamide as a carbon or nitrogen source, but otherwise would have a wild-type phenotype; specifically, they would still be able to use urea as an nitrogen source.

3. Mutation in the more general regulatory genes that are involved in the induced expression of both the acetamidase gene and the two genes for the urea metabolising enzymes, UC and AH. Such mutants would be unable to use either acetamide as a carbon or nitrogen source, and would also be unable to use urea as an N source (i.e. they would require ammonium in the medium).

4. Mutation in the carbon assimilation pathway. Such mutants would not be able to use either acetamide or acetate as a carbon source (i.e. they would be dependent on carbon assimilation through CO_2 fixation and would therefore die when grown in the dark). In these mutants, nitrogen assimilation from urea or acetamide would be normal.

All these mutant classes have been described (Hodson and Gresshoff 1987). Whilst FAM resistance mutants have been generated by spontaneous selection or UV radiation, it should also be possible to generate mutants by insertional mutagenesis using any introduced fragment of DNA. Such mutants should be stable and amenable to gene cloning using the introduced DNA as a handle to isolate genomic DNA flanking the site of insertion.

A strategy for the development of a selectable marker for *Chlamydomonas* would therefore be the isolation of fluoro-acetamide resistant mutants using a gene tagging approach, followed by cloning of the affected gene. Complementation of the mutant with the wild-type gene would then provide a negative selection strategy. Transformants harbouring the gene would be sensitive to FAM.

Acetamide ←——— acetamidase ———→ Ammonium + Acetate
                                    (N source)   (C source)

Urea

urea   allophanate
carboxylase   hydrolase

Ammonium + CO_2

Fluoro-acetamide ——— acetamidase ———→ Fluoro-acetate

Fluoro-citrate
(TCA cycle inhibitor)
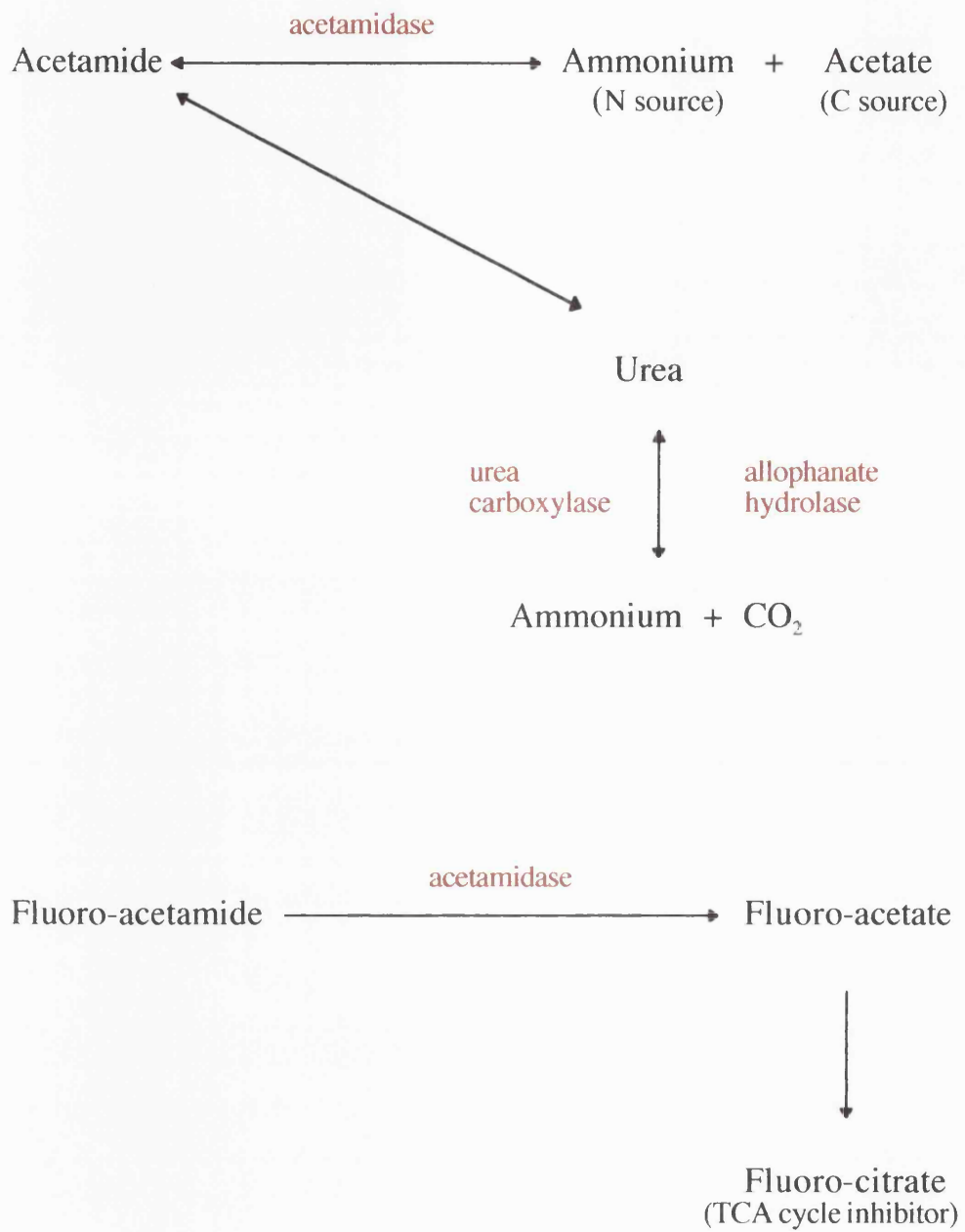
Figure 5.6 *Acetate utilisation pathway in* Chlamydomonas

139

The strategy for development of a negative selectable marker in *Chlamydomonas* is thus:

1. The generation of mutants resistant to FOA or FAM.

2. Further analysis of the mutants to determine at what point in the pathway the mutation arises (e.g. are they *ompd* or *acet* mutants?)

3. Cloning of the affected genes and demonstration that their introduction into the resistance mutant restores sensitivity to the fluoro-analogue.

4. Demonstration of targeted disruption of a *Chlamydomonas* gene using the negative marker in conjunction with a positive marker such as *BLE*.

5.1.2 Antisense down-regulation of genes in *Chlamydomonas*

Antisense RNA is a means of partially or completely repressing the synthesis of a gene product in an attempt to observe steps in a metabolic pathway and to ascertain the relationship between transcript and protein levels. Antisense transformation can result in a dramatic decrease in expression of the homologous host gene (Mol, Blockland et al. 1994) and has provided a powerful tool for establishing the function of plant genes by down-regulation of a specific gene or gene family of interest. Antisense regulation also has commercially important applications in crop improvement (Gasser and Fraley 1989; Fray and Grierson 1993; Bourque 1995). Additionally, antisense inhibition has been observed in the multicellular green alga *Volvox carteri*, a close relative of *Chlamydomonas*. The gene for periplasmic arylsulphatase has been cloned from *Volvox*, and used in the construction of various antisense expression vectors, which were subsequently shown to induce a reduction in arylsulphatase activity in different *Volvox* transformants ranging from 60-79% (Kobl, Scheidlmeier et al. 1996). The exact mechanism of antisense inhibition remains unclear, but the rationale behind the application of antisense genes is the synthesis *in vivo* of complementary RNA, which subsequently hybridises to its target RNA and prevents expression. In eukaryotes the RNA duplex could either be rapidly degraded or could interfere with the normal processing of the mRNA, its transport from the nucleus or translation of the mRNA in the cytoplasm (figure 5.7). Inhibition is restricted to the target gene (and perhaps related genes with high sequence homology). Varying levels of inhibition are seen in different transformants carrying an identical transgene: this leakiness allows for the observation of phenotypic effects of genes whose complete inhibition would be lethal. Such an approach would be extremely useful in *Chlamydomonas* as a means for down-regulation of specific genes without the need for gene targeting.

Target gene   Antisense gene

mRNA   anti mRNA

duplex
formation

mRNA   anti mRNA
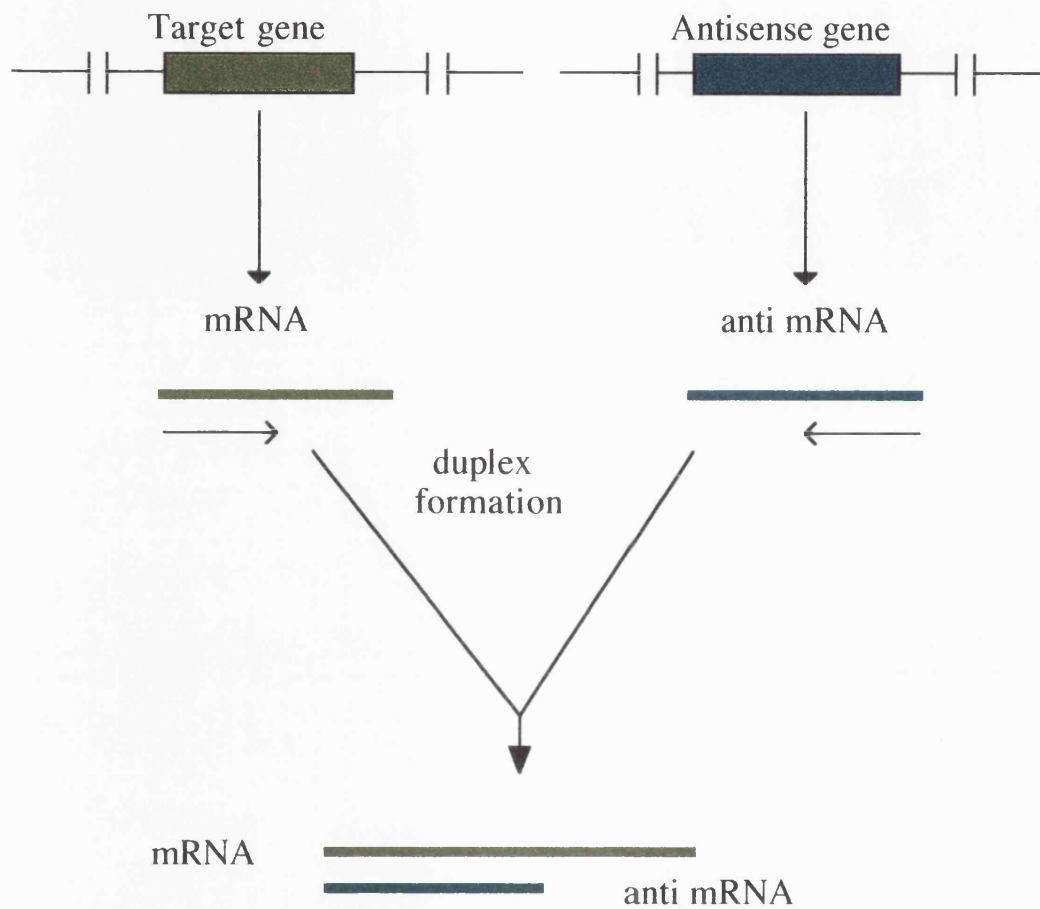
Possible mechanisms of antisense action:

- duplex unstable and degraded
- transport into cytoplasm blocked
- prevention of ribosome binding

Figure 5.7 *Possible mechanisms for antisense down-regulation of genes.*
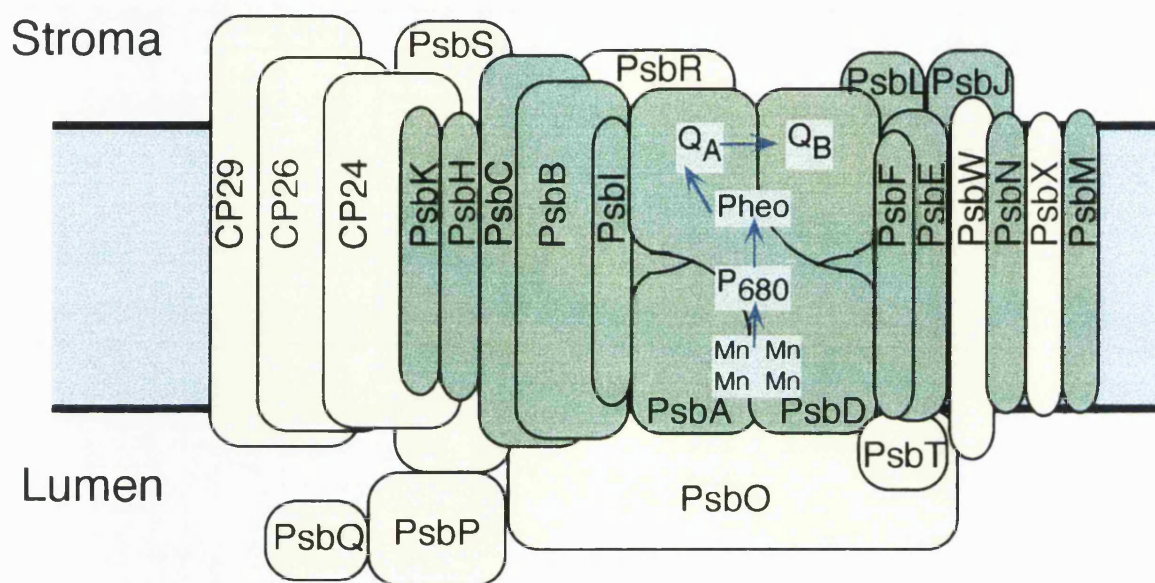
Figure 5.8 *Schematic representation of the photosystem II complex (taken from Robinson and Mant 1997). The chloroplast-encoded subunits are shown in green and the nuclear-encoded subunits are shown in yellow. The polypeptides are shown according to the consensus gene designation for higher plants (psb corresponds to PSII and hence psbH corresponds to PSII-H). The lumenal proteins PsbO, PsbP and PsbQ correspond to the 33, 23 and 16 kDa proteins of the oxygen evolving complex respectively, which are also known as OEE1, OEE2 and OEE3 in* Chlamydomonas.

The single copy nuclear gene *oee1*, encoding the oxygen evolving complex 1 was selected as a candidate for antisense expression in *Chlamydomonas*. The product of *oee1* is the oxygen evolving enhancer protein I (OEE1), a 33 kDa protein that forms part of the oxygen evolving complex of photosystem II (figure 5.7). The OEE1 protein is required for oxygen evolution by the water splitting complex that supplies photosystem II with electrons. The complete gene sequence of *oee1* is known (Mayfield, Schirmer-Rahire et al. 1989) and a non-photosynthetic mutant, FuD44, lacking in synthesis of *oee1* is available. FuD44 carries a transposon insertion in the *oee1* gene (Mayfield, Bennoun et al. 1987, Day and Rochaix 1991). The mutant is deficient in the OEE1 protein and will only grow on a medium containing a reduced carbon source such as acetate. The mutant shows a high fluorescence phenotype due to chlorophyll autofluorescence. The effects of over-expression of *oee1* have also been investigated. No effect on photosynthetic oxygen evolution was observed, though OEE1 protein accumulated to approximately three-fold wild-type levels (Mayfield 1991).

The strategy devised for investigating antisense regulation in *Chlamydomonas* is as follows. An antisense vector was constructed utilising the truncated version of the strong *RBCS2* promoter (see chapter 4) together with a 0.97 kb section of the *oee1* gene in the antisense orientation. The transformants were assayed for any effect on phenotype by fluorescence analysis and for the presence of the transforming DNA by Southern analysis. The positive transformants were then assayed for a decrease in levels of OEE1 protein by western analysis, and for the presence of an antisense transcript by northern analysis.

5.2 Results

5.2.1 Isolation and characterisation of mutants resistant to 5-fluoro-orotic acid

In order to exploit FOA sensitivity as a negative selection in *Chlamydomonas*, it was first necessary to determine the level of sensitivity and the rate of spontaneous resistance to the drug. Wild-type cells were spotted onto plates containing concentrations of 5-fluoro-orotic acid between 1 and 10 mM. A killing concentration of 5-FOA was found to be 4 mM (figure 5.9). Spot tests were also carried out in order to determine the sensitivity of *Chlamydomonas* to 5-fluoro-uracil in order to demonstrate that *Chlamydomonas* is capable of uptake of 5-fluoro-uracil and therefore, presumably, exogenous uracil. Wild-type cells were spotted onto TAP plates supplemented with 5-fluoro-uracil in a range of concentrations between 1 and 10 mM. The cells were found to be sensitive to 5-fluoro-uracil at a concentration of 3 mM (figure 5.10).

cw10   wt   2mM

cw10   wt   3mM

cw10   wt   4mM
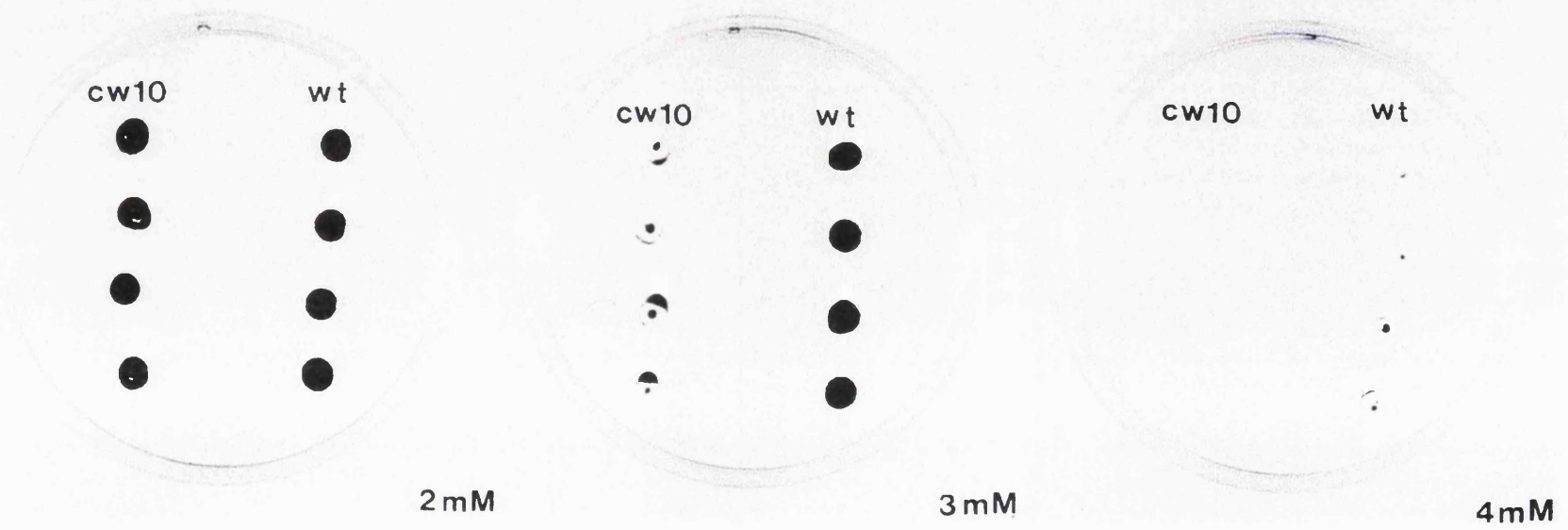
Figure 5.9 *Sensitivity of* Chlamydomonas *to 5-fluoro-orotic acid*
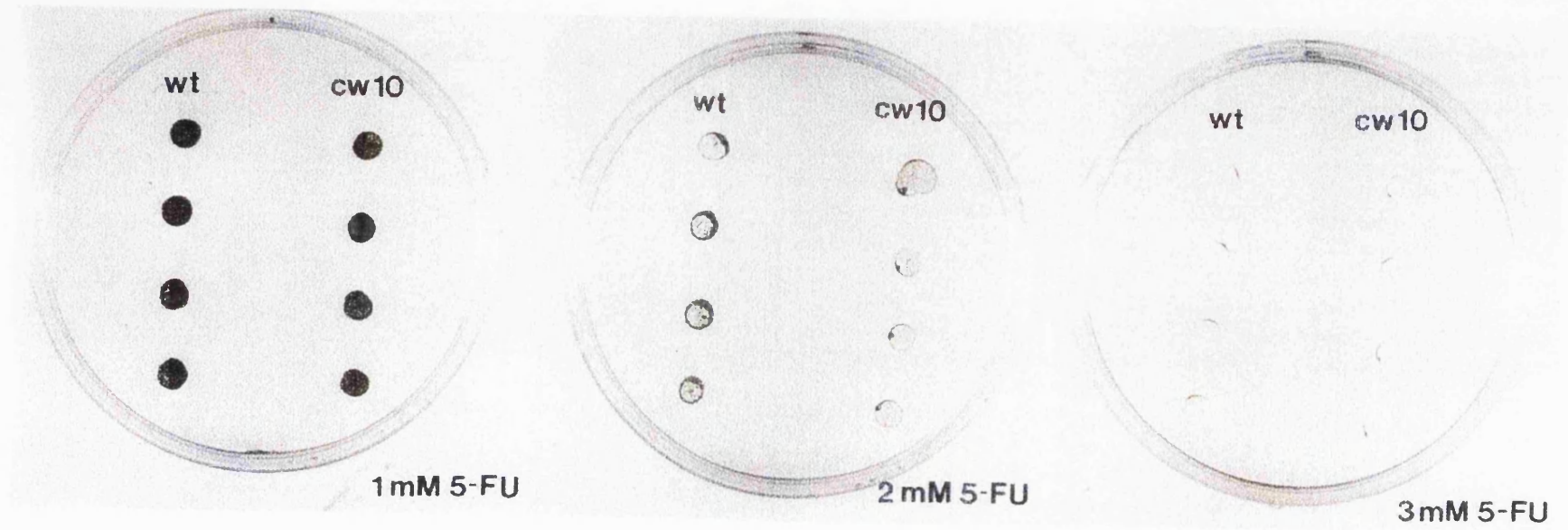
Figure 5.10 *Sensitivity of* Chlamydomonas *to 5-fluoro-uracil*

The level of spontaneous resistance of *Chlamydomonas* to 5-FOA was determined by plating a stationary phase culture (~$2x10^6$ cells per ml) in soft agar onto TAP supplemented with 5 mM 5-FOA and 3 mM uracil. The number of colonies were counted after six days. A low level of spontaneous resistance was found. None of these mutants were found to be uracil-requiring therefore it is doubtful that they were affected in the UMP pathway. A second attempt was made to produce spontaneous mutants using a higher concentration of 5-FOA (10 mM) in the selective medium. Once more, only a low level of spontaneous mutants were recovered. None of the FOA resistant mutants were uracil requiring.

It was decided to try to produce such a mutant using ultra violet radiation as a mutagen. Such an approach is documented as a way of producing nuclear mutations in *Chlamydomonas* (Harris 1989). The method involves exposure of a mid-log phase culture to ultra violet radiation until only 10-15% of the cells remain viable. The cells are then left in darkness for 1 hour to prevent photoreactivation of DNA repair mechanisms, before plating onto selective medium.

In order to determine the exposure required to produce a 10% survival rate, a liquid culture at mid-log phase was irradiated using a UV source for 15 minutes. At 30 second intervals a sample was removed and plated onto TAP in soft agar after a 1 hour recovery period in the dark. After six days the number of colonies for each plate were counted and plotted against the exposure time to produce a kill curve (figure 5.11). From this graph, the UV exposure required to produce a 10% viable culture could be determined. A second mid-log phase culture was prepared and subject to UV radiation for five minutes and 30 seconds (as determined from the kill curve) with constant stirring. The culture was then plated onto TAP supplemented with 5 mM 5-FOA and 3 mM uracil. This UV mutagenesis method produced a high level of resistance mutants. 20 such mutants were selected for further analysis, but none were found to be uracil requiring (figure 5.12).

Given that the UMP biosynthesis pathway is known be essential in most other organisms, and the fact that FOA resistant, uracil requiring mutants have been isolated from diatoms, yeast and *Arabidopsis*, it is unlikely that the UMP biosynthesis pathway does not exist in *Chlamydomonas*. One possibility is that the mutants isolated so far are uptake mutants i.e. they have an alteration in their membrane structure so that they are unable to assimilate 5-FOA, and therefore display a resistant phenotype. If this is the case, then it is likely that if sufficient mutants are screened, a 5-FOA resistant, uracil requiring mutant will eventually be isolated. A second possibility is that the mutants isolated so far are affected in one of the steps of the UMP biosynthesis pathway, but the mutants are not uracil requiring as there is a second pathway through which they are able to synthesise uracil.

UV mutagenesis curve

10% survival

no. colonies

Time (min)

no. colonies
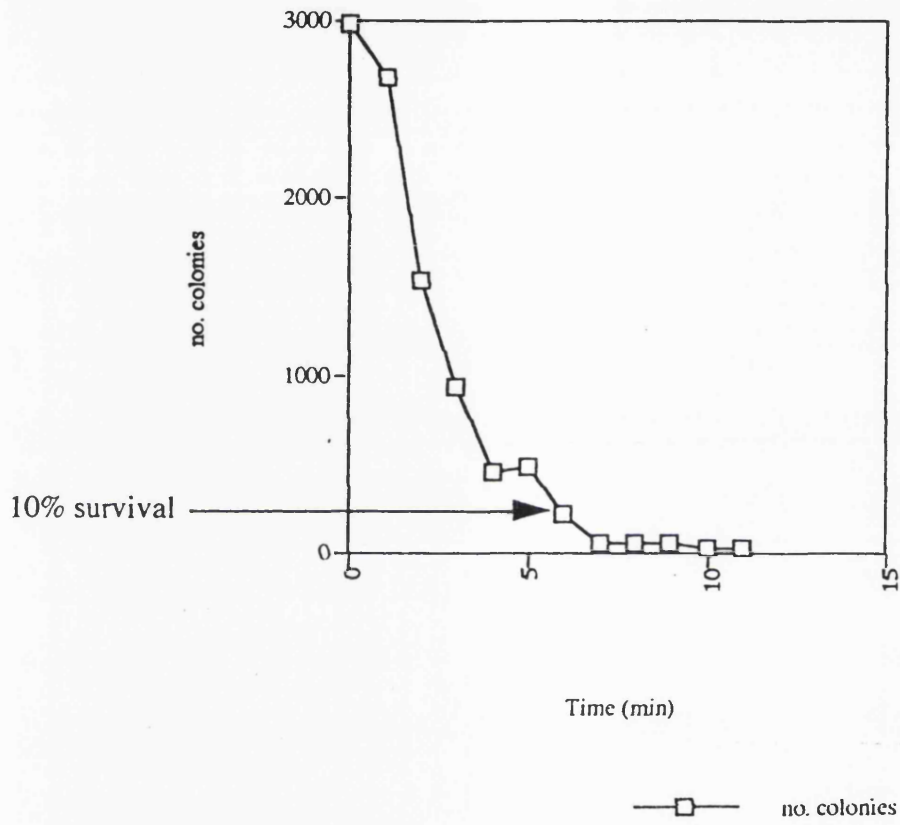
Figure 5.11 Chlamydomonas *'kill' curve for UV mutagenesis*

147

TAP   TAP + 5mM uracil   TAP + 4 mM 5-FOA
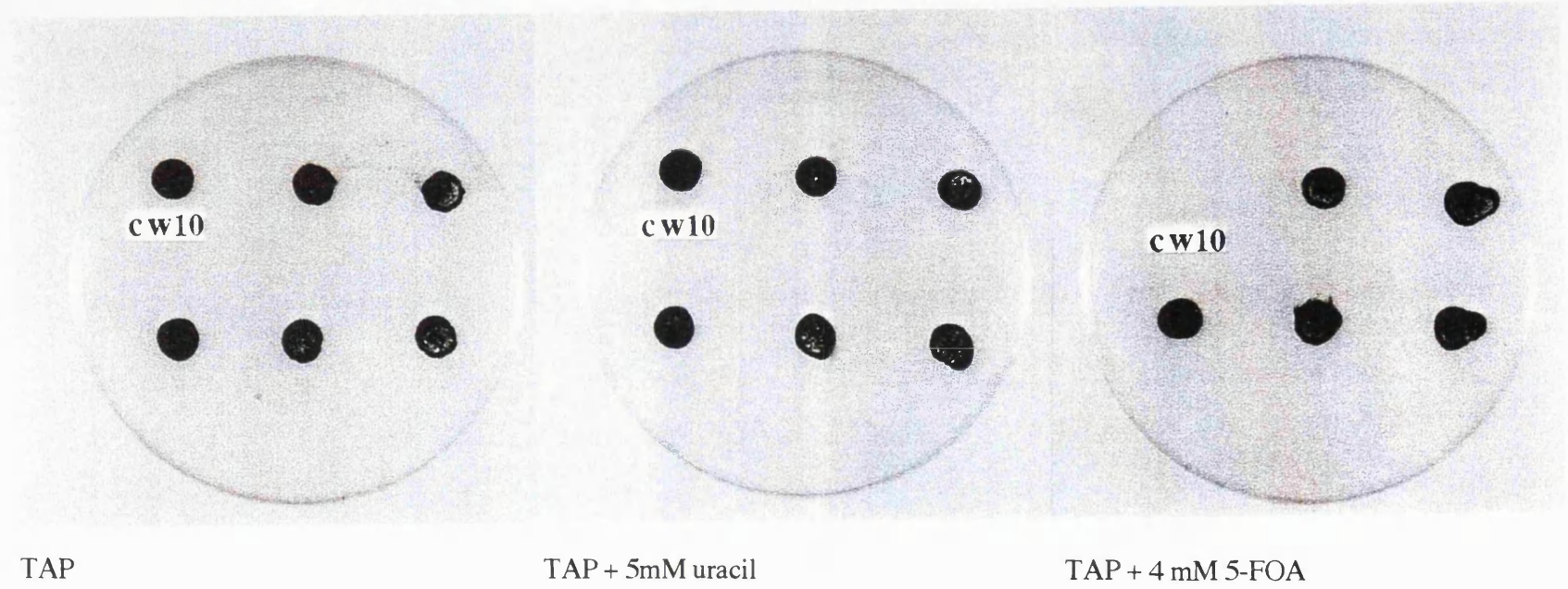
Figure 5.12 *Mutants of* Chlamydomonas *resistant to 5-FOA*

5.2.2 Cloning of the wild-type gene encoding *Chlamydomonas* OMPD

Attempts were made to clone the wild-type OMPD gene from *Chlamydomonas*. The gene encoding orotidine-5'-monophosphate decarboxylase is well conserved across species (Benito, Diaz Minguez et al. 1992). An alignment of the amino acid sequences for OMPD was obtained from the protein sequence databases using a multiple sequence alignment program (Lasergene by DNAStar). NB. the gene encoding OMPD is known as *pyr4*, *pyr*G, *pyr*F, *pyr5-6*, *ura3* and *ura4* in various species. A particularly well conserved sequence of nine amino acid residues was selected. An oligo was raised against this sequence using the *Chlamydomonas* preferred codon usage (figure 5.13). Two degenerate codons were introduced into the oligo sequence to account for degeneracy in the *Chlamydomonas* genetic code. The best conserved sequence of amino acids was:

F L I F E D R

The sequence of the oligo, written 5' to 3' was:

ompd oligo: **TTC CCS ATY TTC GAG GAY CG**

where S=G or C and Y=C or T

This *ompd* consensus oligo was used as a primer firstly to try to amplify the wild-type *ompd* gene by PCR from a *Chlamydomonas* cDNA library, and then if this PCR approach failed, to screen the library using the radio-labelled oligo as a probe. A *Chlamydomonas* cDNA library made from wild-type cells of *Chlamydomonas* CC137 (mt⁺) (donated by Professor J V Moroney at Louisiana State University) was used to screen for *ompd* in *Chlamydomonas*. The library was constructed in the lambda Uni-ZAP XR vector (Stratagene), pre-digested with *Eco*RI and *Xho*I. A feature of the lambda Uni-ZAP range of vectors is that cloning of cDNAs is unidirectional i.e. clones insert into the library vector in one orientation only. The other primer for the PCR reaction was raised against the T7 promoter region of the lambda ZAP vector. The sequence of the extended T7 primer, written 5' to 3' was:

T7 ext; GTA ATA CGA CTC ACT ATA GGG CGA ATT GGG

Various PCR reaction conditions, times and temperatures were used in order to amplify the *ompd* gene from the library, but these attempts did not prove successful, and no PCR product was obtained.

The oligo was also used to search for *ompd* in the library by hybridisation. *E. coli* C600 was transfected with the library and plated out. The resulting plaques were transferred to nitro-cellulose. The *ompd* oligo was end-labelled and used to probe the nitro-cellulose

Consensus sequence of *ompd* (from Blast search):

```
                  ┌──────────────────────────────────────────────────┐
                  │ ▓▓░░▓░▓▓▓░░░░░░░░░░░░▓░░░░▓▓▓░▓▓░▓▓▓▓░▓▓▓░▓▓░▓░▓▓▓▓░░▓▓ │
                  LNALAQKHNFLIFEDRKFADIGNTVKLQYHGGVYRIAEWAHITNCHIVPGEGI
                   90        100       110       120       130       140
s.pombe ura4     LVALGKKHRFLIFEDRKFADIGNTVKLQYASGVYKIASWAHITNCHTVPGEGI   98
phycomyces py    LEALAKKHDFLIFEDRKFADIGNTVKHQYEKGVYKIASWSHITNAHTVPGEGI   98
p.chrysogenum    LNALAQKHNFLIFEDRKFIDIGNTVQKQYHNGTLRISEWAHIINCSILPGEGI  101
k.lactis ura3    LKALAEKYKFLIFEDRKFADIGNTVKLQYTSGVYRIAEWSDITNAHGVTGAGI   96
t.reesei ura3    LASLARKHGFLIFEDRKFGDIGHTVELQYTGGSARIIDWAHIVNVNMVPGKAS  104
p.chrysogenum    LNALAQKHNFLIFEDRKFIDIGNTVQKQYHNGTLRISEWAHIINCSILPGEGI  101
A.niger pyrG     LKALAQKHNFLIFEDRKFIDIGNTVQKQYHRGTLRISEWAHIINCSILPGEGI   98
Candida B URA    LLALAKKYNFMIFEDRKFADIGNTVKSQYKGGVYQIAKWSDITNAHGVTGSGI  100
```

↓

F   L   I   F   E   D   R

design of oligo using
Chlamydomonas
best codon usage

↓

TTC   CCS   ATY   TTC   GAG   GAY   CG

where S=G or C and Y=C or T

Figure 5.13 *Design of oligo to consensus protein sequence of* ompd

filters. Eight positive clones were isolated which were subsequently proved to be false positives by further analysis.

The failure to clone *Chlamydomonas ompd* from the cDNA library could be due to one of two factors:

1. The degenerate *ompd* oligo is not specific enough to hybridise to its target for hybridisation studies or amplification by PCR.

2. The gene is not represented in the particular cDNA library used. A different cDNA library may contain the *Chlamydomonas* gene encoding OMPD. Screening of several cDNA libraries may be required.

5.2.3 Isolation and characterisation of mutants resistant to fluoro-acetamide

Previous work with *Chlamydomonas* has determined that the organism is sensitive to a 150 μM concentration of fluoro-acetamide on solid growth media (Hodson and Gresshoff 1987). The level of spontaneous resistance of *Chlamydomonas* to fluoro-acetamide was determined by scoring on plates lacking in ammonium, but containing 5 mM urea, 150 μM FAM and acetate. The plates were grown in continuous light using a wild-type strain of *Chlamydomonas*, CC137 (mt$^+$). Hodson and Gresshoff reported a spontaneous rate of mutation of 1 in 10^{-4}-10^{-5} viable plated cells, equating to several hundred colonies per plate. This is a high level of background mutation, which would make scoring for insertional mutants extremely difficult. It was decided to carry out the screen using acetate-free plates to lower the rate of spontaneous mutation by increasing the stringency of selection for mutants affected in acetate utilisation. This should give rise to an acceptable background level of spontaneous mutation, such that identification of colonies arising from insertional mutagenesis might be possible.

To test this theory, growth medium lacking in ammonium and acetate was prepared (Tris min-N). Growth medium lacking in ammonium but containing acetate (TAP-N) was prepared as a control. Tris min-N and TAP-N growth media were made in the same way as Tris min and TAP (see Materials and Methods section, Chapter 2) using Beijerink salts lacking in ammonium chloride.

5.2.4 Determining the spontaneous resistance frequency to FAM in *Chlamydomonas*

A cell wall-less mutant of *Chlamydomonas*, CC-849 (mt$^+$) was grown to early stationary phase ($\sim$2x10^6 cells per ml) and resuspended in Tris min-N. This cell suspension was then

	TAP-N	Tris min -N
no. colonies	37	3
	303	6
	70	12
	100	6
	25	12
Average per plate	171	8

Table 5 a   *Comparison of numbers of colonies obtained
by spontaneous resistance on Tris min - N and
TAP - N*

plated onto selection plates plus and minus acetate (TAP-N and Tris min-N supplemented with 5 mM urea and 150 μM FAM). The colonies were counted after 6-10 days (Table 5 a). The rate of spontaneous resistance of *Chlamydomonas* grown on selective media with and without acetate was calculated:

Rate of spontaneous resistance = number of mutants isolated
                                 total number of cells

Total number of cells in 0.1 ml of cell suspension = 5×10^6

∴ rate of spontaneous resistance of cells grown on selective media with acetate:

$$= \frac{171}{5 \times 10^6} = 3.42 \times 10^{-5}$$

and rate of spontaneous resistance of cells grown on selective media without acetate:

$$= \frac{8}{5 \times 10^6} = 1.6 \times 10^{-6}$$

The rate of spontaneous mutation obtained for cells grown on selective media without acetate (Tris min -N) was decreased by a factor of 10 relative to that obtained for cells grown on selective media containing acetate (TAP-N). The latter was similar to that obtained by Hodgson and Gresshof in their original studies (Hodson and Gresshoff 1987). A rate of spontaneous mutation of 1.6×10^{-6} for those cells grown without acetate was deemed an acceptable level of background mutation for isolation of mutants defective in acetamidase. Further analysis of twelve FAM spontaneous resistance mutants indicated that three of the mutants were unable to grow in the dark. Such mutants are defective in the carbon assimilation pathway and are not able to use either acetamide or acetate as a carbon source (i.e. they would are dependent on C assimilation through CO_2 fixation) in accordance with the mutant classes isolated by Hodgson and Gresshof. If the genes affected in this class of mutants could be isolated, then the genes would be valuable as both positive (restoration of growth in the dark) and negative (restoration of FAM sensitivity) selectable markers.

5.2.5 Isolation of mutants resistant to FAM by insertional mutagenesis

One strategy for isolating such genes is insertional mutagenesis. The lac-α cassette, generated in the Purton laboratory, was used as an insertional mutagen. The cassette was
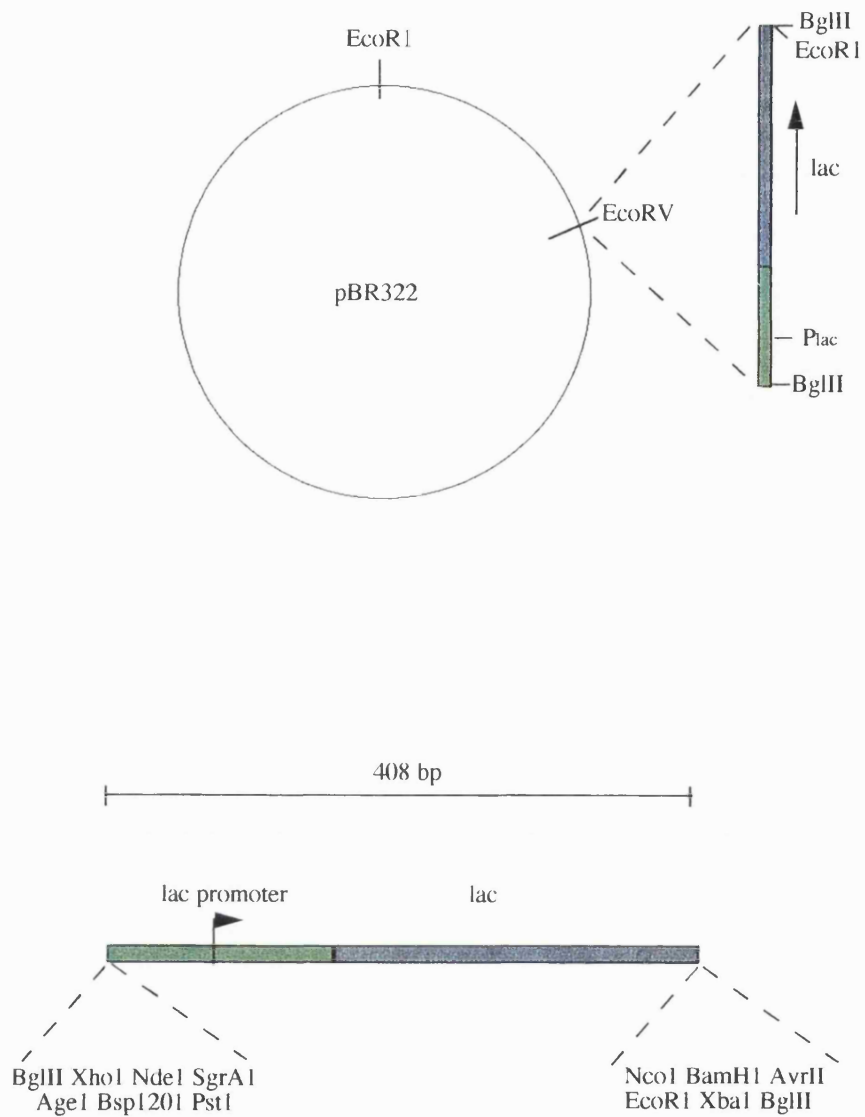
Figure 5.14 *The lac-α cassette*

made by PCR amplification of the lac-α promoter and coding region of the *lacZ* gene of *E. coli*. The PCR product was cloned into a 'non-lac' vector (pBR322) and transformed into the *E. coli lac⁻* host DH5-α (figure 5.14). Blue/white selection was employed to select transformants that had gained *lac* activity. Use of lac-α as an insertional mutagen provides a gene tagging approach with the additional advantage that cloned fragments disrupted by the lac-α cassette may be selected in *E. coli* by means of blue/white selection which provides a simplified screen for the tagged DNA. The 408 bp segment encoding lac-α was excised from the plasmid using *Bgl*II and resuspended at a concentration of 1 mg/ml. This DNA was then used as an insertional mutagen by transformation of the *Chlamydomonas* strain CC-849 using the glass bead transformation method (Kindle 1990). The resulting transformants were plated onto selective medium, and the numbers of colonies counted after six days (Table 5 b). Isolation of FAM resistant mutants resulting from an insertion of the lac-α DNA should represent valuable material for cloning of the affected genes. The rate of spontaneous mutation versus the rate of mutation obtained by insertional mutagenesis was calculated:

Rate of spontaneous resistance = $\dfrac{\text{number of mutants isolated}}{\text{total number of cells}}$

Total number of cells in 0.1 ml of cell suspension = 5×10^6

∴ rate of spontaneous resistance of cells

$$= \frac{133}{5 \times 10^6} = 2.66 \times 10^{-6}$$

and rate of mutants isolated by insertional mutagenesis

$$= \frac{297}{5 \times 10^6} = 5.94 \times 10^{-5}$$

The increase in the rate of mutation using insertional mutagenesis versus the rate of spontaneous mutation indicates that the method provides an efficient method for isolation of mutants defective in uptake of acetamide. The acetate utilisation pathway represents a high number of targets (about 50 acetate utilisation genes are known) and the frequency of

	Insertional mutagenesis	Spontaneous resistance
No. colonies	350	152
	259	247
	326	163
	326	266
	252	199
	-	200
	-	294
	-	140
	-	141
	-	125
Average	297	133

Table 5 b   *Comparison of numbers of colonies obtained by insertional mutagenesis and by spontaneous resistance*

156

mutations obtained with this method is comparable to that observed with conventional mutagenic agents like UV, EMS and MNNG (Adam 1993). The main interest in insertional mutagenesis is that isolating a mutant in this way provides a tool for cloning the corresponding gene (Gumpel and Purton 1994). An ever-increasing list of genes have been cloned using this approach, including those required for flagellar function (Tam and Lefebvre 1993), sulphur metabolism (Davies et al. 1996), nitrogen metabolism (Zhang and Lefebvre 1997) and chloroplast gene expression (Gumpel, Ralley et al. 1995). The isolation of FAM resistant mutants in this manner means that some potentially represent lac-α tagged mutants.

The next step in isolation of the tagged genes defective in such mutants is Southern analysis of the FAM mutants to confirm the presence of lac-α transforming DNA in the mutant genome. To isolate the sequences flanking a putative lac-α tagged gene, total cellular DNA should be digested with a suitable restriction enzyme, and size fractionated on an agarose gel. DNA fragments from the region of the gel containing the lac-α transforming DNA can then be purified, ligated into a 'non lac' E. coli plasmid vector, and transformed into a E. coli lac⁻ host. Positive colonies can be selected by blue/white selection and the genomic DNA flanking the lac-α DNA excised from its E. coli vector and used to screen a Chlamydomonas genomic library for the wild-type gene (Gumpel, Ralley et al. 1995). Once isolated the wild-type gene could be used in complementation of the original FAM mutant to provide a selection strategy. Unfortunately, the isolation of FAM requiring mutants occurred near to the end of the experimental part of this thesis, and time constraints meant that these final experiments were not possible.

5.2.6 Construction of the antisense expression vector poeeα

In order to investigate whether the antisense down-regulation of genes is feasible in Chlamydomonas, an antisense plasmid was constructed utilising a fragment of the gene encoding oxygen evolving enhancer protein, oee1. A 0.97 kb HpaI fragment of oee1 was isolated from a plasmid encoding wild-type oee1 cDNA and cloned into the multiple cloning site of pSP105 cut with HincII (figure 5.15). The orientation of the insert was determined by restriction analysis using BamHI and PstI. Two positive transformants were selected that had the 0.97 kb oee1 insert in the reverse orientation. The clone was sequenced across the ATG start codon to confirm the orientation of the oee1 insert. The promoter of the construct was then truncated from 740 to 180 bp which has been shown to increase the strength of the promoter (Lumbreras, Stevens et al. 1998). This was achieved through excision of a 560 bp

```
   1 ccgcgcgcgc aagcaccgat cgggctgtac ggctagcgag atcaatttcc gtaccgcgtc
  61 ggggcaggca tgaagcgcgt tgtttctctt acggagttgt cccccgtcaa ccgctagtcg
 121 accatttctc ctttgcacac cacagcagca aagatggccc tccgcgctgc ccagtccgct
 181 aaggccggtg tccgtgccgc ccggcccaac cgtgccaccg ccgtggtgtg caaggcgcag
 241 aaggtcggcc aggccgccgc tgccgctgct ctggccaccg ccatggttgc cgggtcggcc
 301 aacgccctga ccttcgacga gattcagggc ctgacctacc tgcaggtcaa gggctccggc
 361 attgccaaca cctgccccgt gctggagagc ggcaccacca acctgaagga gctgaaggcc
 421 ggctcctaca agctggagaa cttctgcatt gagcccacct ctttcaccgt gaaggaggag
 481 agccagttca agggtggcga gactgagttt gtcaagacca agctcatgac ccgcctgacc
 541 tacaccctgg acgccatgtc cggctcgttc aaggtcggct ctgacggctc cgctgagctg
 601 aaggaggatg acggcattga ctacgccgcc accaccgtcc agctgcccgg tggcgagcgc
 661 gtggctttcc tgttcaccat caagcagttc gatggcaagg gcaccctgga cggcatcaag
 721 ggtgacttcc tggtgccctc gtaccgcggc tcgtccttcc tggaccccaa gggccgcggt
 781 ggctcgaccg gctacgacaa cgccgtggct ctgcccgccc gcgctgatgc cgaggagctg
 841 ctgaaggaga acgtgaagat caccaaggcc ctgaagggct ctgccgtgtt ctccgttgcc
 901 aaggtggacc ccgtgaccgg cgagatcgct ggtgtgttcg agtccatcca gccctcggac
 961 actgacctgg gcgccaagcc ccccaaggac atcaaggtca ccggcctgtg gtacgcccag
1021 ctgaagtaag cgctttcgcc ccccgaagat aaacacttct ggaggaggag gacgctactg
1081 cttagcatgg tcgaccgctg cgaggaggag acgtgagcca ggatgctgga ggagctctag
1141 cgggtgtaca tcgtaggtgc tgcttccggg attggaaggg taggcaaagc agtgctgcgg
1201 gcgggcagac ggcgtgtcaa aacgctcgcc tgtgtgcggt gcagtgctgg agcgtttttg
1261 ttgaggacgc tcaaactcgg gtggattcag aggtcggggc gcggaggttg cgtggttgat
1321 cgagcggggg ccaaacgttg ccgctgccgc acgcctgccg tatgcgcgcc cggcaacagc
1381 cgtggtgctg tcggagcgcg agtgccggtt ttggaggtct aggagtgcta ggttggtgcg
1441 agcgccgtgt catccggctt ttcgagaaga aagatacagt gtgcccaaaa gtgtgtgatt
1501 attgtgtttc atagtggctc tcgaaacaag gccgctttcc cagtgtagcc tgacacttcg
1561 gagccctatt tccccgactc gcttctcgtc gtgcatgcgc tttgctggca gtagtgcagc
1621 accgctgctc tcagtcaata acaggctggc aaaggccctg tgcacagatt agtcatgagg
1681 cgtgtttgcg gcggcgtgta acacccaaga tgag
```

Figure 5.15 *cDNA sequence of* oee1 *in* Chlamydomonas *with 0.97 kb* HpaI *fragment highlighted in blue. This section was excised from a vector encoding the wild-type* oee1 *cDNA and cloned into a* Chlamydomonas *expression vector in the antisense orientation.*
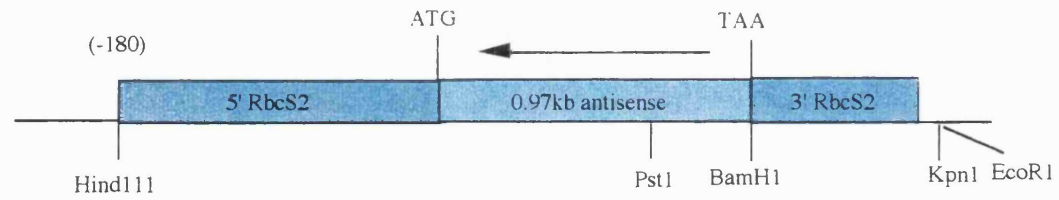
Figure 5.16   *Antisense expression vector for* Chlamydomonas *incorporating a 0.97 kb fragment of the single copy nuclear gene* oee1

SmaI/HindIII promoter fragment from the construct, followed by religation to create the plasmid p*oeeα* (figure 5.16).

p*oeeα* was transformed into *Chlamydomonas* by co-transformation with the plasmid pARG7.8 encoding argininosuccinate lyase (Purton and Rochaix 1995) using the glass bead transformation method (Kindle 1990). The recipient strain for transformation was the cell-wall deficient, arginine requiring strain CC-363 (mt⁻). Selection of transformants was by restoration of arginine-independent growth.

5.2.7 Fluorescence analysis of antisense transformants

Non-photosynthetic *oee1* mutants of *Chlamydomonas* show a highly fluorescent phenotype when excited by light at <620 nm due to chlorophyll autofluorescence (Mayfield, Bennoun et al. 1987). Antisense transformants in liquid culture were assayed for high fluorescence compared to wild-type cells by fluorescence analysis using a Perkin Elmer spectrofluorimeter and on solid media using fluorescence video imaging. The non-photosynthetic *oee1* mutant FuD44 was used as a positive control. The cells were grown in dim light. The concentration of cells in liquid culture was measured by chlorophyll assay, and adjusted so that each measurement was taken for the same cell concentration. FuD44 was shown to exhibit high fluorescence when excited by light at 620 nm compared with wild-type cells (figure 5.17), however the *oee* antisense construct had no discernible effect on the fluorescence state of the cells. This was also found to be the case using fluorescence video imaging. As this initial screen for highly fluorescent antisense transformants was unsuccessful, several antisense transformants were selected for Southern analysis to confirm the presence of the antisense construct.

5.2.8 Southern analysis of antisense transformants

Total genomic DNA preparations were made for eleven of the transformants and digested with *Pvu*II. *Pvu*II cuts *oee1* cDNA at positions 642 bp and 1020 bp, which lie within the *Hinc*II antisense fragment. The Southern shows the presence of genomic *oee1* as two bands present at 4100 bp and 750 bp. Positive transformants show the presence of bands at 1000 bp and 378 bp (figure 5.18). These transformants were then subject to western analysis to look for a decrease in the level of OEE1 protein.

5.2.9 Western analysis of antisense transformants

As high fluorescence of the transformants was not apparent, the five positive transformants were further screened by western analysis for a decreased amount of OEE1 protein using equal amounts of protein extract. Wild-type total cellular protein was used as a positive control and FuD44 total cellular protein as the negative control. The resulting western blot
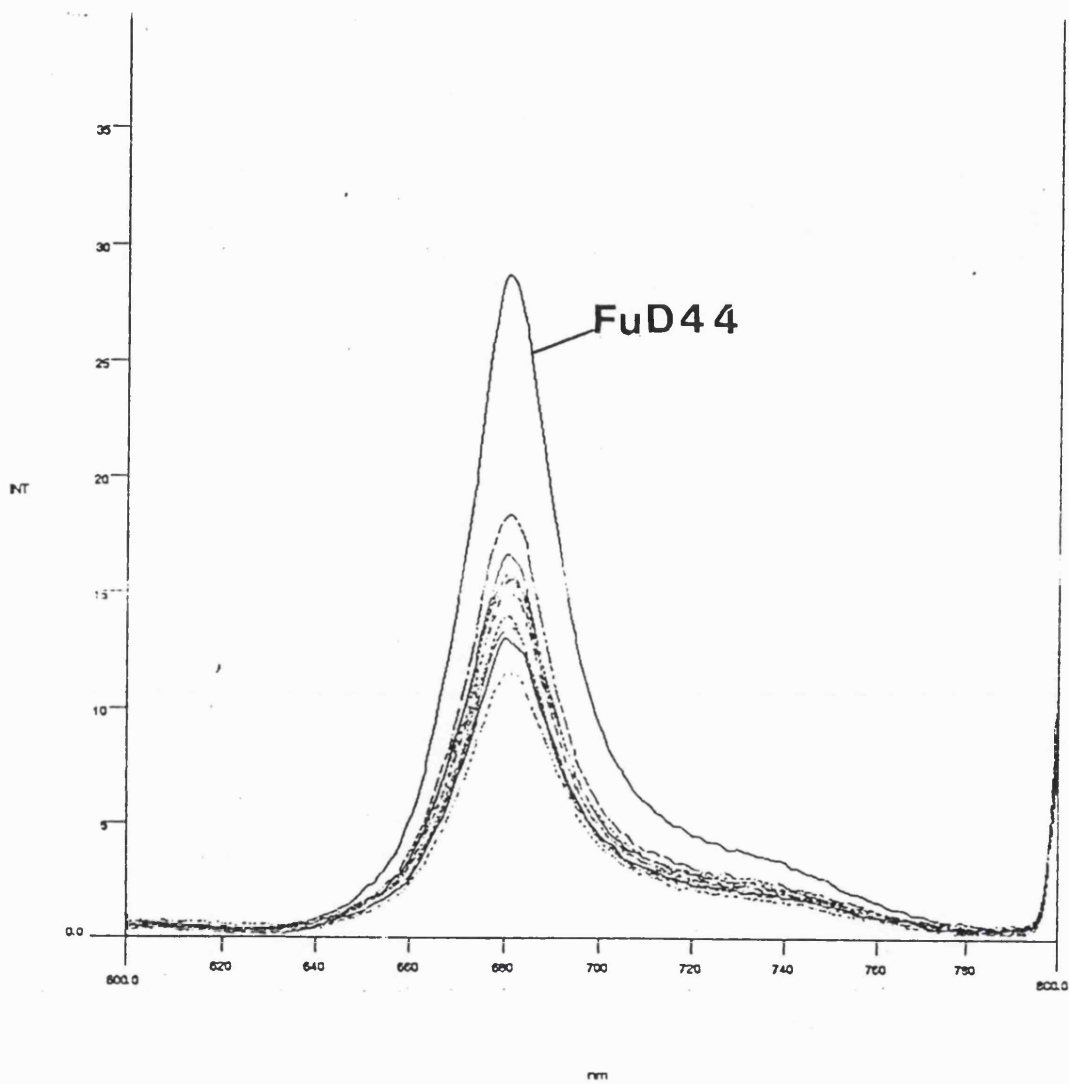
Figure 5.17 *Fluorescence analysis of* oeeα *transformants*

161
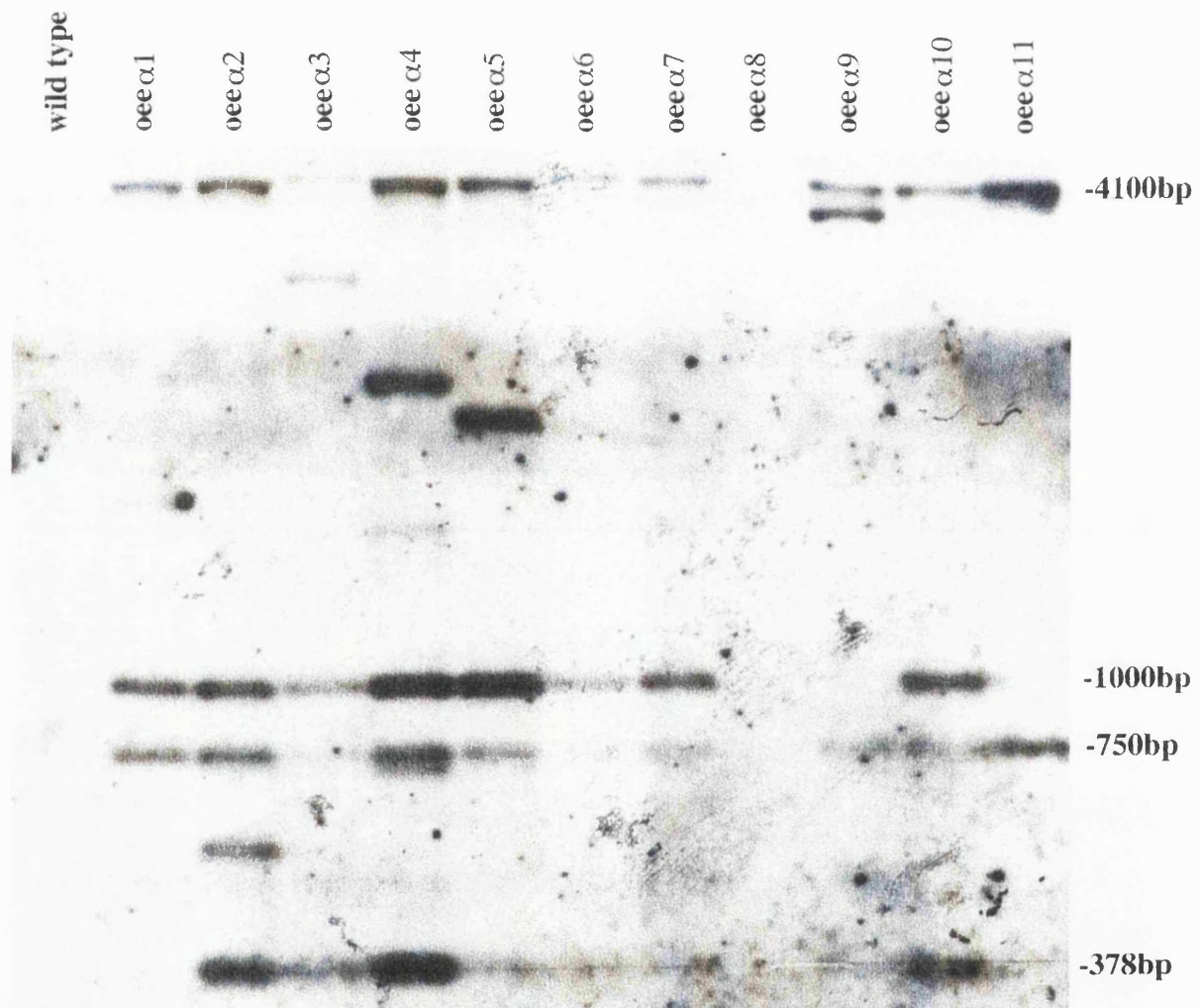
Figure 5.18 *Southern analysis of poeeα transformants. 10 μg of total genomic DNA extracted from the poeeα transformants and from wild-type cells were digested with PvuII and separated on a 1% agarose gel by electrophoresis. The DNA was transferred to a nylon filter by Southern blotting and probed with a 0.92 kb fragment of ^{32}P labelled oee1 cDNA DNA. The filtered was washed for 2 x 20 minutes in 0.1% SDS, 2x SSC at room temperature, then for 2x 20 minutes in 0.1% SDS, 0.2 x SSC at 65 °C. The filter was then autoradiographed for 24 hours at -70 °C.*
The figure shows the presence of genomic oee1 as two bands present at 4100 bp and 750 bp. Please note that the reason the same binding is not seen in the wild-type track is that the wild-type DNA had apparently been degraded and was seen as a smear on the agarose gel. Positive transformants show the presence of bands at 1000 bp and 378 bp.
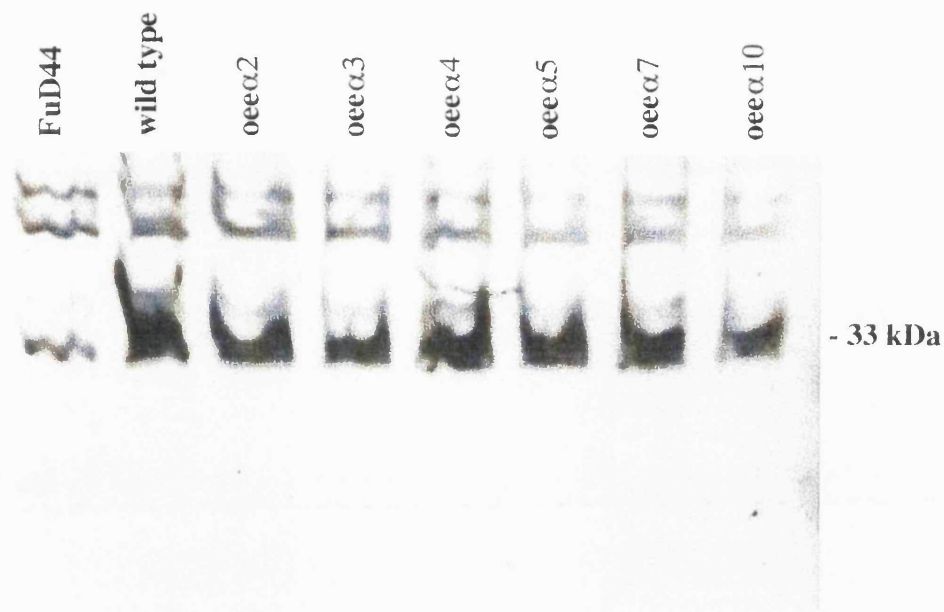
Figure 5.19 *Western analysis of poeeα transformants. Total protein extracts (measured on an*
equal chlorophyll basis equivalent to 10 μg of chlorophyll) from the positive poeeα transformants
were loaded onto a tris-tricine gel. The same quantity of protein extracted from FuD44, a mutant
with a disrupted copy of the wild-type oee1 gene was used as a negative control, and total protein
extract from wild type cells was used as a positive control. The proteins were separated by gel
electrophoresis and transferred to a nitrocellulose filter. The resulting blot was probed with an
antibody raised against p oee1 protein from pea. Primary antibody binding was detected using a
horse radish peroxidase labelled secondary antibody and subsequently autoradiography.
Note that there is no significant decrease in the amount of OEE1 protein detected relative to wild-
type. Some OEE protein is also seen in the Fud44 null mutant. This could be due to 'spill-over'
during gel loading, or to the fact that reversion of the mutant population may have occurred during
maintenance on agar plates.
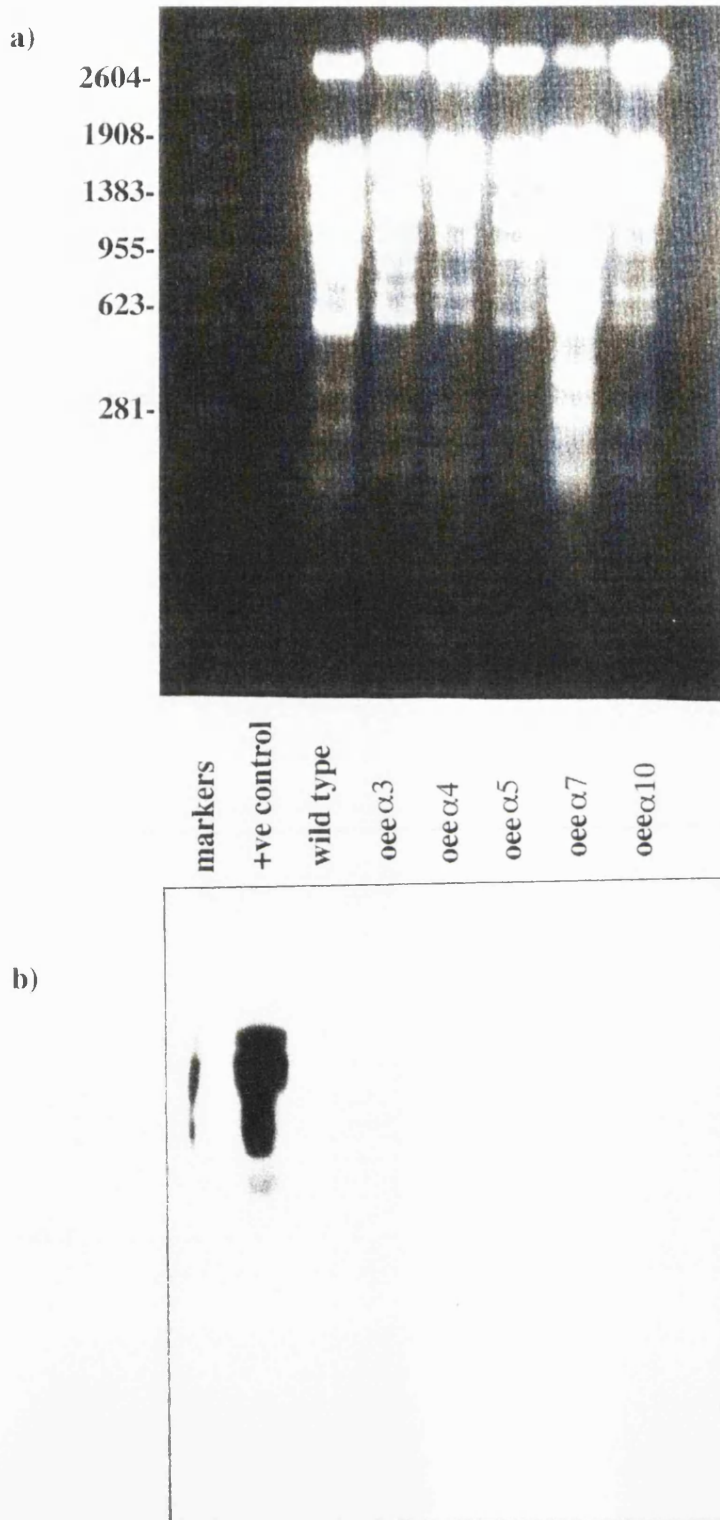
Figure 5.20 *Northern analysis of positive poeeα transformants showing a) RNA gel probed using a 0.92 kb fragment of* ^{32}P *labelled oee cDNA, and b) resulting autorad. Total RNA from WT and poeeα transformants was separated on a 0.5% denaturing agarose gel, and transferred to a nitrocellulose membrane. Since a positive signal cannot be detected for the genomic oee1 gene despite abundant amounts of RNA present on the gel, this experiment cannot be interpreted as negative for the presence of an antisense transcript .*

was probed with poly-clonal antibodies raised against pea OEE1 protein (donated by Professor John Gray, University of Cambridge). The amount of OEE1 protein was found to be constant in all of the transformants assayed relative to FuD44 which showed a significant decrease in the amount of OEE1 protein (figure 5.19).

This evidence suggests that although the antisense DNA may be transcribed, it has no discernible effects on OEE1 protein levels. It would therefore be logical to assume that antisense downregulation of the *oee1* gene does not take place in *Chlamydomonas*. However evidence from tobacco has shown that antisense down-regulation of the *oee1* gene does not affect the corresponding protein levels, indicating that the mRNA levels for this component is not rate limiting for protein accumulation and that severely reduced amounts of these transcripts still allow normal plant development (Palomares, Herrmann et al. 1993). For this reason northern analysis was carried out in order to determine whether the antisense construct is transcribed in the positive transformants, and whether transcription of the antisense fragment has a discernible effect on wild-type mRNA levels.

5.2.10 Northern analysis of antisense transformants

The transformants were subject to northern analysis to assay for the presence of an antisense transcript. No antisense transcript was detected in the antisense transformants. The northern analysis also failed to detect mRNA from the single-copy nuclear gene encoding *oee1*, although the positive DNA control was detected (figure 5.20). This indicates that the DNA probe is not sensitive enough to detect the presence of an antisense transcript, and the results were therefore inconclusive. Greater sensitivity could be achieved through use of a radiolabelled RNA probe, synthesised by transcription *in vitro* from the poeeα plasmid, however time constraints meant that it was not possible to perform this experiment.

5.3 Discussion

5.3.1 Development of a negative selectable marker for *Chlamydomonas*

Although the strategies outlined above for the development of a negative selectable marker for *Chlamydomonas* have failed to yield a suitable marker gene in the short term, given a longer time scale it is possible that both strategies could yield a suitable candidate.

Presuming that *Chlamydomonas* does not have an alternative pathway for uracil biosynthesis that allows it to do without its UMP biosynthetic pathway, then the FOA resistant mutants isolated so far must be FOA assimilation mutants. In this case, screening of sufficient colonies should eventually yield a mutant affected in the UMP biosynthetic pathway that is both 5-FOA resistant and uracil-requiring. Isolation of wild-type *Chlamydomonas ompd* may be possible using a different cDNA library, or by RACE (Rapid Amplification of cDNA Ends) from *Chlamydomonas* total mRNA (Frohman 1990).

The RACE protocol generates cDNA by using PCR to amplify copies of the region between a single point in the transcipt and the 3' or 5' end. To use the RACE protocol, one must raise a primer to a short section of the DNA sequence and a primer that anneals to the natural 5' end or 3' end or a synthetic poly-A tail.

The insertional mutagenesis strategy for isolation of FAM resistant mutants was successful in that the approach has been shown to yield many hundreds of transformants with the required phenotype. Analysis of these mutants by Southern analysis using lac-α DNA as a probe should show which mutants have become resistant to FAM due to insertion of lac-α DNA to produce a suitable mutant background for genomic complementation. The acetamidase utilisation pathway is complex and there are several candidate genes other than acetamidase that will yield a FAM resistant phenotype. Cloning of the wild-type gene responsible for the mutant phenotype is simplified by the fact that the mutant version of the gene should be tagged with lac-α DNA, and with a greater number of target genes for isolation of the expected phenotype this approach is more likely to provide a means of negative selection than isolation of FOA resistant mutants.

5.3.2 Antisense down-regulation of genes in *Chlamydomonas*

Despite the success of antisense as a means for down-regulation of genes in higher plants and in *Volvox carteri* (Kobl, Scheidlmeier et al. 1996) successful antisense down-regulation of genes in *Chlamydomonas* has proven elusive. Other researchers have encountered similar problems in attempts to down-regulate *Chlamydomonas* genes by antisense [Karen Kindle - personal communication]. The inability to express antisense DNA in *Chlamydomonas* therefore appears to be intrinsic to the organism. Lack of evidence of an antisense transcript would suggest that antisense RNA is recognised as foreign by the organism, since it is not translatable and may lack stabilising secondary structure, and as such is rapidly degraded. This also appears to be the case with *gfp* mRNA (see Chapter 4 discussion). One strategy for avoiding such rapid degradation of the antisense transcript would be to reduce the size of the antisense transcript. Presumably, RNases require a binding site to the RNA in question before they can exert their degradative activity, so a smaller antisense transcript is less likely to be subject to rapid turnover. A small antisense RNA could be raised against the RNA binding site around the AUG initiation codon and would exert its effect by preventing translation of the mRNA. The efficiency of this method would depend on the mechanism of antisense inhibition being due to interference by the antisense RNA with the normal processing of the mRNA at the level of its translation in the cytoplasm.

The strategies outlined in this chapter deal with methods for altering the expression patterns of genes at the DNA level, by gene targeting to knockout specific genes, and at the RNA level by targeting the mRNA by antisense. Another strategy might be to try to knockout

genes at the functional protein level. Over-expression of a gene encoding a mutant non-functional version of a protein could be used as a way of creating a mutant phenotype for a given protein. The effects would be similar to those achieved by antisense down-regulation or gene knockout by altering the kinetics of wild-type protein production. The protein could be modified to include extra amino acid residues by site directed mutagenesis that would make it non-functional.

Chapter 6
CONCLUDING REMARKS AND DISCUSSION

Chapter 6: Concluding remarks and discussion

6.1 Précis - how far have the goals of this thesis been realised?

As explained in the introduction to this thesis, further procedures for molecular genetic investigation are required in order to exploit fully the potential of *Chlamydomonas*. For example, it is desirable to develop a detailed molecular map of the *Chlamydomonas* nuclear genome. Also there is a need to find a suitable reporter molecule that can act as a marker for gene expression and protein localisation. Other beneficial advances would include a way of targeting specific genes by selection for rare homologous recombination events and a means for down-regulation of genes by the application of antisense technology. This thesis presents the results of investigation into several possible new tools for nuclear transformation.

6.1.1 Analysis of ESTs from *Chlamydomonas*

Firstly, the results of a pilot study of Expressed Sequence Tag analysis for *Chlamydomonas* were presented. A robust and routine method was developed for the isolation of ESTs from *Chlamydomonas* by PCR and sixteen cDNAs chosen at random from a *Chlamydomonas* cDNA library were sequenced from both ends. These sequences represent *Chlamydomonas* ESTs, and were analysed for sequence similarities to other organisms and for functional motifs. Of the sixteen ESTs, two clones were shown to have exact sequence homology with well characterised nuclear genes encoding ferredoxin and ribulose-5'-bisphosphate carboxylase. A further four clones were identified that showed similarity to previously characterised *Chlamydomonas* genes and were putatively identified as genes encoding new members of *Chlamydomonas* multigene families e.g. cabII (see chapter 3). Three of the remaining clones encoded products with some significant homology to proteins from other organisms such as cytochrome p450 and a negatively phytochrome regulated protein, whilst the remaining clones failed to produce a significant match with any known protein. Whilst these latter two classes clearly represent new *Chlamydomonas* genes, it is impossible to confidently assign a function to them. Preliminary EST analysis of a further 84 clones produced by the EST PCR protocol further demonstrated the effectiveness of the procedure and produced some interesting results. In total, 100 clones were sequenced and were found to include copies of previously characterised *Chlamydomonas* genes *RBCS2* and *cabII*, as well as the genes encoding *Chlamydomonas* ferredoxin and an ADP-ribosylation factor (Mamon et al. 1995). The BLAST searches also indicated the presence of number of other interesting clones in the library including one encoding a soluble starch synthase. In addition, several clones encoding a product that is remarkably similar to PsaL or subunit V of photosystem I, from a number of organisms. PsaL has not so far been cloned from *Chlamydomonas*. Also included were a ribosomal 60S protein that showed high homology

to the 60S ribosomal protein from *Arabidopsis* and yeast, and another 60S acidic ribosomal protein that showed similarity to the 60S acidic ribosomal protein P2 from *Trypanosoma cruzi*.

6.1.2 The use of green fluorescent protein as a reporter for *Chlamydomonas*

The results of the investigation of the use of Green Fluorescent Protein as a reporter molecule for *Chlamydomonas* have proved disappointing. It does not appear to be possible to routinely express *gfp* in *Chlamydomonas*, a problem that appears to be intrinsic to the codon usage of the alga. The DNA does not appear to be transcribed or translated, although the lack of evidence for a *gfp* transcript could be due to the fact that the mRNA is recognised as foreign to *Chlamydomonas*, and as such is rapidly degraded. It seems that the major problem for the expression of the *A. victoria* gene in *Chlamydomonas* is one of codon usage. If this is indeed the case, extensive changes to the codon usage by construction of a synthetic gene will be required if *gfp* is to be expressed in the alga. Such an approach has been tried but was not found to be successful [Peter Hegemann-personal communication]. This is most probably due to the fact that changing the codon usage of a gene is an unknown quantity in terms of introduction of secondary structure.

Furthermore such an approach is expensive and time consuming. With the continued emergence of an ever increasing range of mutant versions of GFP it is possible that a *gfp* variant will eventually emerge that is suitable for expression in the alga. It seemed that the GFP variant EGFP might provide such an alternative, however there appears to be a further problem with expression of this variant. Lack of expression of EGFP could perhaps be due to the existence of a cryptic splice site within the coding sequence of EGFP. Investigation into the mechanism of intron splicing in *Chlamydomonas* has not lead to the resolution of a clear consensus splice sequence. Identification of such an element within the coding regions of a transgene may need to wait until better knowledge is available of the factors regulating RNA splicing in the alga.

6.1.3 Gene targeting and antisense down-regulation of nuclear genes in *Chlamydomonas*

An attempt was made to develop a negative selectable marker for homologous transformation based on the creation of mutants defective in uracil or acetate utilisation and their restoration to wild-type phenotype by transformation with the wild-type gene. Two pathways were investigated in which substitution of a substrate for its fluoro-analogue causes cell death in *Chlamydomonas;* that for UMP biosynthesis, and that for acetate utilisation. Mutants resistant to 5-FOA were created by UV mutagenesis but were subsequently found to be defective in some aspect of uracil utilisation other than the gene encoding OMPD. Furthermore, attempts to clone wild-type *ompd* from *Chlamydomonas*

were not successful and the approach was subsequently halted. Mutants affected in acetate utilisation were also produced, this time by insertional mutagenesis, which provides a tag on the affected gene. Unfortunately, due to time constraints it was not possible to fully characterise the mutants and clone the affected gene by virtue of the fact that it is tagged with an insertional mutagen. On the whole, however, this insertional mutagenesis approach has produced some encouraging results. Clearly, fluoro-analogue resistance is a promising approach for the isolation of mutants affected in some aspect of substrate metabolism. There are, of course, many alternative fluoro-analogues of metabolic substrates that could be used to isolate a *Chlamydomonas* mutant defective in a single enzyme of a metabolic pathway, and cloning of corresponding gene should be relatively straightforward if the mutant is produced by insertional mutagenesis. The main disadvantage of this approach is that complementation of a mutant is not always desirable as a means of selection. In order to use this method of selection with a given mutant strain, it is first necessary to cross the strain to the fluoro-analogue resistant mutant background. Also, this means of negative selection is leaky - there will be a significant background of non-homologous transformants in which the negative selectable marker has failed to integrate, has been subject to rearrangements or mutations during integration, or is poorly expressed in the transformant. However, such a system could still provide a useful means for enriching for homologous recombination events.

Finally the use of antisense expression was investigated. An antisense section of the *Chlamydomonas* nuclear gene encoding *oee1* was introduced into the nuclear genome under the control of a strong *Chlamydomonas* promoter and the resulting transformants analysed for phenotypic effects and for the presence of an antisense transcript.* No effect on phenotype could be observed, and an antisense transcript could not be detected in the transformants. This leads to speculation that the lack of antisense inhibition in *Chlamydomonas* could, like *gfp* expression, be due to a failure in translation ability. The antisense transcript is recognised as foreign to the organism and as such rapidly turned over by the cell. One prospect for antisense expression in *Chlamydomonas* could be to use small mRNAs which will be less susceptible to degradation due to the fact that RNase binding to its target RNA will not be possible, or to use an antisense RNA targeted to the ribosome binding site around the ATG start codon.

6.2 Prospects for *Chlamydomonas* molecular techniques
6.2.1 EST analysis and the *Chlamydomonas* genome sequencing project

The pilot study for EST analysis in *Chlamydomonas* has produced a practical, robust method for the generation and analysis of ESTs that, given the current interest in the establishment of a *Chlamydomonas* Genome Sequencing Project, should prove invaluable to *Chlamydomonas* genome researchers. As detailed in Chapter 3, one of the preliminary

---

*However, lack of detection of wild-type OEE RNA in the Northern analysis means that this experiment cannot be taken as proof of lack of antisense expression.

steps for the *Chlamydomonas* genome project is the development of an Expressed Sequence Tag database to identify genes expressed in *Chlamydomonas*. The cDNAs sequenced could be isolated from cells grown under different conditions. These ESTs could be used as molecular markers in the building up of a physical map of the genome, and provide information on the expressed gene content of cells grown under varying conditions. An approach that is currently proving extremely fruitful in the popular model organism *Arabidopsis* (for which a genome mapping project is now well underway) is the establishment of collection of mutants generated by transposon or T-DNA tagging (McKinney, Ali et al. 1995). The mutants are stored in seedbanks in pooled lots (e.g. 100). DNA prepared from these mutant pools has been made available for researchers to screen by PCR for mutations in genes of interest. This protocol is dependant on the ability to routinely amplify sequence from a disrupted gene using primers specific to the gene of interest and to the T-DNA tag when the genomic DNA is diluted from 99 other transformants. The advantage of using this approach with *Arabidopsis* is that the collections of mutants can be stored in seedbanks, and the mutant pools can easily be regenerated. Insertional mutagenesis could be used to produce such a mutant library for *Chlamydomonas*, and the mutants pooled into manageable size lots. This protocol is dependant on the ability to routinely amplify sequence from a disrupted gene using primers specific to the gene of interest and to the insertional mutagen, as with *Arabidopsis*. The process is also dependant on the development of freezing protocols that are efficient enough to allow the freezing of pools of transformants. Maintenance of individual transformants on agar would be extremely labour intensive and maintenance of the stocks as pools on agar may lead to selection against certain insertional mutants within the population. It may be some time before this approach can successfully be applied to *Chlamydomonas*, but as freezing technology improves for *Chlamydomonas*, it is not unreasonable to expect that the technology for routinely freezing stocks of mutants will shortly emerge.

6.2.2 The development of a reporter molecule for *Chlamydomonas*

Reporter technology for *Chlamydomonas* is still at a somewhat primitive stage. Although reporters for promoter activity (e.g. *ARS* and *RSP3* - see chapter 4) have been successfully used in the alga, their usefulness for a reporter of promoter activity is somewhat limited. What is needed is something, that like GFP can be visualised and used as a reporter of gene expression or a protein tag in living cells. A variety of such reporter molecules have emerged for other organisms over the last decade. For example the GUS reporter system, encoded by the bacterial *uidA* gene, which is capable of utilising a chromogenic substrate to produce a blue colorant, and has been successfully used in plants such that it is now routine (Jefferson 1989). However, the use of GUS in *Chlamydomonas* is far from routine. Although GUS has successfully ₐbeen used in the chloroplast of *Chlamydomonas* (Sakamoto et

172

al. 1994), many attempts to utilise GUS as a marker for nuclear gene expression and protein localisation have failed in the alga. The problem again appears to be one of codon usage, and successful expression of GUS will probably only be achieved by the engineering of a codon optimised gene with all its pitfalls and expense.

The future for a reporter for *Chlamydomonas* may not be the use of green fluorescent protein, but may lie in the use of a luciferase reporter molecule. Of the many different strategies available for using as a genetic reporter, the luciferases offer an ideal solution. Reporter measurements are nearly instantaneous and they are exceptionally sensitive and quantitative. Many different organisms ranging from bacteria and fungi to fireflies and fish are endowed with the ability to emit light, but the genes encoding for the luciferase protein are not homologous, and the luciferin substrates are also different, falling into unrelated chemical classes. Biochemically, all known luciferases are oxygenases that utilise molecular oxygen to oxidise a substrate luciferin (luciferin; literally the 'light bearing' molecule) with the formation of the product molecule in an electronically excited state (Hastings 1996). Genes encoding the luciferases and luciferins are not conserved across species. Thus luciferase and luciferin are generic terms and the organism from which they are obtained must be specified. Firefly luciferase is by far the most commonly used of the bioluminescent reporters, and has been expressed successfully in transgenic plants (Ow, Wood et al. 1986). Transient expression of firefly luciferase has also been attained in the protoplasts of the green alga *Chlorella ellipsoidea* (Jarvis and Brown 1991). A report in the form of a conference abstract for the forthcoming *Chlamydomonas* VIIIth international meeting of the use of the coral luciferase as a vital marker for gene expression in the chloroplast of *Chlamydomonas* has also recently been published on the World Wide Web (Min'ko, Holloway et al. 1998). If the problem of heterologous gene expression in *Chlamydomonas* is, as it appears, one of codon usage then it is just possible that a dinoflagellate luciferase might provide the answer. A luciferase from the marine alga *Gonyaulax polydra* has been cloned, sequenced and analysed (Bae and Hastings 1994). Its luciferase is a single chain with an estimated mass of 130 kDa and its light emitting reaction involves the enzymatic oxidation of its tetrapyrole luciferin by molecular oxygen. The luciferase protein was expressed in *E. coli,* and the coding sequence produced an active luciferase. The deduced amino acid sequence and codon usage indicate that this gene is strongly biased to G or C residues in the third position of each codon, and the overall G+C content of the coding region is 61.6 %, with 78.9 % of codons with G or C as their third positions. Thus the codon usage is remarkably similar to that of *Chlamydomonas* (Rochaix 1995). and it may be feasible to express this luciferase in *Chlamydomonas*. The main drawback is that there is at present no commercially available source of the substrate luciferin - the user would need to extract the luciferin from the marine alga or clone and express the biosynthetic genes for substrate from *E. coli* for herself, a time-consuming business!

173

6.2.3 The development of a negative selectable marker for targeted gene disruption in *Chlamydomonas*

An alternative strategy for a 'positive-negative' selection strategy has recently been demonstrated in *Dictyostelium*, that may provide an alternative means of negative selection for *Chlamydomonas* (Morrison, Marschalek et al. 1997). A lethal suppressor tRNA, tRNA^{Glu}II(UUA), was used as a simple, negative selectable marker to enrich for gene disruption. All attempts to introduce an active copy of the tRNA^{Glu}II(UUA) ochre suppressor into *Dictyostelium* have failed (Dingermann, Brechner et al. 1989). This tRNA is believed to be lethal to *Dictyostelium* because 97% of all known *Dictyostelium* genes use UAA as their termination codon due to an extremely high A+T rich genome. A disruption construct was made utilising *neo* (encoding neomycin phosphotransferase) as a positive marker flanked by regions of DNA homologous to the *ecmA* gene locus, which has a low frequency of homologous recombination in *Dictyostelium*. *Neo* confers resistance to G418 (genticin). The construct also carried the suppressor tRNA gene as a negative marker. The construct was found to reduce the transformation frequency 20-fold compared to transformation with a construct that was identical bar the tRNA suppressor gene. This is indicative of a reduction in the background of non-homologous integration. A similar tRNA suppressor approach may be possible in other organisms with biased nucleotide composition. In *Chlamydomonas*, the majority of genes also terminate with the UAA stop codon, although a tRNA gene has not yet been isolated. However, the nature of tRNAs is such that they are less subject to variation across species and it may be possible to use the tRNA^{Glu}II(UUA) ochre suppressor from *Dictyostelium* in *Chlamydomonas*. This approach has a three fold advantage over the fluoro-analogue resistance approach: firstly there is no need to used a mutant background; secondly, the tRNA is a small so that the background of non-homologous transformants in which the negative selectable marker has failed to integrate or has been subject to rearrangements or mutations during integration is likely to be reduced; thirdly there is no need to use a toxic substrate such as fluoro-acetamide for selection.

6.3 Concluding remarks

Although the 'green yeast' *Chlamydomonas reinhardtii* has a way to go in terms of molecular techniques, the prospects for the organism are bright. The fact that *Chlamydomonas* performs functions that other organisms do not such as photosynthesis and motility has already lead to profound insights into these cellular processes. Reporter gene technology and a means of gene targeting for *Chlamydomonas* will eventually come as new technology emerges, given enough resources. Furthermore the *Chlamydomonas* Genome Sequencing Project, coupled with a complete EST database and clone collection,

promise to make *Chlamydomonas* a model to rival other popular models such as yeast, *E. coli* , *Arabidopsis* and *Dictyostelium*.

REFERENCES

References

Adam, M., K. E. Lentz, et al. (1993). "Insertional mutagenesis to isolate acetate requring mutants in *Chlamydomonas reinhardtii*." *FEMS Microbiol. Lett.* **110**: 265-268.

Adams, M. D. (1992). "Sequence identification of 2,375 human brain genes." *Nature* **355**: 632-634.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." *J Mol. Biol.* **215**: 403-410.

Antequera, F. and A. P. Bird (1988). "Unmethylated CpG islands associated with genes in higher plant DNA." *EMBO J.* **7**: 2295-2299.

Bae, Y. M. and J. W. Hastings (1994). "Cloning, sequencing and expression of dinoflagellate luciferase DNA from a marine alga, *Gonyaulax polydra*." *Biochim. Biophys. Acta* **1219**: 449-456.

Bassi, R., S. Y. Soen, et al. (1992). "Characterisation of chlorophyll a/b proteins of photosystem I from *Chlamydomonas reinhardtii*." *J. Biol. Chem.* **267**: 25714-25721.

Bassi, R. and F.-A. Wollman (1991). "The chlorophyll-a/b proteins of photosystem II in *Chlamydomonas reinhardtii*: Isolation, characterisation and immunological cross-reactivity to higher-plant polypeptides." *Planta* **183**: 423-433.

Benito, E. P., J. M. Diaz Minguez, et al. (1992). "Cloning and sequence analysis of the *Mucor circinelloides* pyrG gene encoding orotidine-5'-monophosphate decarboxylase: use of pyrG for homologous transformation." *Gene* **116**(1): 59-67.

Bennetzen, J. L. and B. D. Hall (1982). "Codon selection in yeast." *J. Biol. Chem.* **257**: 3026-3031.

Birnboim, B. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." *Nucl. Acids Res.* **7**: 1513-1523.

Blowers, A. D., L. Bogorad, et al. (1989). "Studies on *Chlamydomonas* chloroplast transformation: foreign DNA can be stably maintained in the chromosome." *Plant Cell* **1**(1): 123-32.

Boeke, J. D., F. Lacroute, et al. (1984). "A positive selection for mutants lacking orotidine-5'-monophopsphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance." *Mol. Gen. Genet.* **197**: 345-346.

Boer, P., L. Bonen, et al. (1985). "Genes for respiratory chain proteins and ribosomal RNAs are present on a 16 kilobase-pair DNA species from *Chlamydomonas reinhardtii* mitochondria." *Proc. Natl. Acad. Sci. USA* **82**: 3340-3044.

Boer, P. and M. W. Gray (1988). "Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA." *EMBO J.* **7**: 3501-3508.

Booher, R., C. E. Alta, et al. (1989). "The fission yeast cdc2/cdc13/suc1 protein kinase: regulation of catalytic activity and nuclear localisation." *Cell* **58**: 485-497.

Bourque, J. E. (1995). "Antisense strategies for genetic manipulation in plants." *Plant Sci.* **105**: 125-149.

Boxer, S. G. (1996). "Structural biology. Another green revolution." *Nature* **383** (6600): 484-485.

Boynton, J. E., N. W. Gillham, et al. (1988). "Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles." *Science* **240**: 1534-1538.

Boynton, J. E. and N. W. Gillham (1993). "Chloroplast transformation in *Chlamydomonas*." *Methods in Enzymology* **217**: 510-538.

Brand, A. (1995). "GFP in *Drosophila*." *Trends Genet.* **11**(8): 324-325.

Brown, L. E., S. L. Sprecher, et al. (1991). "Introduction of exogenous DNA into *Chlamydomonas reinhardtii* by electroporation." *Mol. Cell. Biol.* **11**: 2328-2332.

Capecchi, M. R. (1989). "Altering the genome by homologous recombination." *Science* **244** (4910): 1288-1292.

Casper, S. J. and C. A. Holt (1996). "Expression of the green fluorescent protein-encoding gene from a tabacco mosaic virus-based vector." *Gene* **173**: 69-73.

Cerutti, H., A. M. Johnson, et al. (1997). "A eubacterial gene conferring spectinomycin resistance on *Chlamydomonas reinhardtii*: integration into the nuclear genome and gene expression." *Genetics* **145**(1): 97-110.

Chalfie, M., Y. Tu, et al. (1994). "Green fluorescent protein as a marker for gene expression." *Science* **263**: 802-805.

Chen, C. Y. and M. Inouye (1990). "Supression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons *E. coli* genes." *Nuc. Acids Res.* **18**: 1465-1473.

Chiu, W., Y. Niwa, et al. (1996). "Engineered GFP as a vital reporter in plants." *Curr. Biol.* **6**(3): 325-30

Cody, C. W., D. C. Prasher, et al. (1993). "Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein." *J. Biochem.* **32**: 1212-1218.

Cormack, B. P., R. H. Valdivia, et al. (1996). "FACS-optimised mutants of the green fluorescent protein (GFP)." *Gene* **173**: 33-38.

Cubitt, A. B., R. Heim, et al. (1995). "Understanding, improving and using green fluorescent proteins." *Trends Biochem.* **20**: 448-455.

Dangeard, P.-A. (1888). "Recherches sur les algues inferieures." *Ann. Sci. Nat.* **7**(4): 105-175.

Davies, J. P., D. P. Weeks, et al. (1992). "Expression of the arylsulfatase gene from the beta 2-tubulin promoter in *Chlamydomonas reinhardtii*." *Nuc. Acids Res.* **20**(12): 2959-2965.

Crameri, A., E. A. Whitehorn, et al. (1996). "Improved green fluorescent protein by Molecular Evolution using DNA shuffling." Nature Biotechnology **1 4**: 315-319.

Day, A. and J. D. Rochaix (1991). "Conservation in structure of TOC1 transposons from *Chlamydomonas reinhardtii.*" *Gene* **104**(2): 235-239.

Davies, J. P., F Yildiz F., et al. (1996). "Sac1, a putative regulator that is critical for survival of *Chlamydomonas reinhardtii* during sulphur deprivation." *EMBO J.* **15:** 21520-21529.

Day, A. and J. D. Rochaix (1991). "A transposon with an unusual LTR arrangement from *Chlamydomonas reinhardtii* contains an internal tandem array of 76 bp repeats." *Nuc. Acids Res.* **19**(6): 1259-66

de Hostos, E. L., J. Schilling, et al. (1989). "Structure and expression of the gene encoding the periplasmic arylsulfatase of *Chlamydomonas reinhardtii.*" *Mol. Gen. Genet.* **218**(2): 229-39.

Debuchy, R., S. Purton, et al. (1989). "The argininosuccinate lyase gene of *Chlamydomonas reinhardtii*: an important tool for nuclear transformation and for correlating the genetic and molecular maps of the ARG7 locus." *EMBO J.* **8**(10): 2803-2809.

Delagrave, S., R. E. Hawtin, et al. (1995). "Red-shifted excitation mutants of the green fluorescent protein." *BioTechnology* **13:** 151-154.

Diener, D. R., A. M. Curry, et al. (1990). "Rescue of a paralyzed-flagella mutant of *Chlamydomonas* by transformation." *Proc. Natl. Acad. Sci. USA* **87**(15): 5739-5743

Dietmaier, W. and S. Fabry (1994). "Analysis of the introns in genes encoding small G proteins." *Curr. Genet.* **26**(5-6): 497-505.

Dingermann, T., T. Brechner, et al. (1989). "Non-sense suppression in *Dictyostelium discoideum.*" *Devel. Genet.* **11:** 410-417.

Dopf, J. and T. M. Horiagon (1996). "Deletion mapping of the *Aequorea victoria* green fluorescent protein." *Gene* **173:** 39-44.

Dumas, P., M. Bergdoll, et al. (1994). "Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering." *EMBO J.* **13:** 2483-2492.

Dunahay, T. G. (1993). "Transformation of *Chlamydomonas reinhardtii* with silicon carbide whiskers." *BioTechniques* **15**(3): 452-454.

Dunahay, T. G., E. E. Jarvis, et al. (1996). "Genetic engineering of microalgae for fuel production." *App. Biochem. Biotech.* **34/35:** 331-339.

Ecker, J. R. and R. W. Davis (1986). "Inhibition of gene expression in plant cells by expression of antisense RNA." *Proc. Natl. Acad. Sci. USA* **83:** 5372-5376.

Ehrig, T., D. J. O'Kane, et al. (1995). "Green fluorescent protein mutants with altered fluorescent excitation spectra." *FEBS Letts.* **367:** 163-166.

Epel, B. L., H. S. Padgett, et al. (1996). "Plant virus movement protein dynamics probed with a GFP-protein fusion." *Gene* **173:** 75-79.

Erickson, J. M. (1996). Chapter 32. "Chloroplast transformation: current results and future prospects." *Oxygenic Photosynthesis: The Light Reactions.* J. M. Erickson. Netherlands, Kluwer Academic Publishers: 589-619.

Fernandez, E., R. Schnell, et al. (1989). "Isolation and characterisation of the nitrate reductase structural gene of *Chlamydomonas reinhardtii." Proc. Natl. Acad. Sci. USA* **86**: 6449-6453.

Fray, R. G. and D. Grierson (1993). "Molecular genetics of tomato fruit ripening." *Trends Genet.* **9**(12): 438-443.

Frohman, M. A. (1990). "RACE: Rapid Amplification of cDNA Ends." *PCR Protocols: A guide to Methods and Applications,* Academic Press Inc: 28-38.

Funke, R. P., J. L. Kovar, et al. (1997). "Intracellular carbonic anhydrase is essential to photosynthesis in *Chlamydomonas reinhardtii* at atmospheric levels of CO_2: Demonstration via genomic complementation of the high CO_2 requiring mutant *ca-1[1]." Plant Physiol.* **114**: 237-44.

Gasser, C. S. and R. T. Fraley (1989). "Genetically engineered plants for crop improvement." *Science* **244**: 1293-1298.

Goffeau, A. (1996). "Life with 6000 genes." *Science* **274**: 546-567.

Goldschmidt-Clermont, M. (1986). "Sequence, evolution and differential expression of the two genes encoding variant small subunits of ribulosebiphosphate-carboxylase/oxygenase in *Chlamydomonas reinhardtii." J. Mol. Biol.* **191**: 421-432.

Goldschmidt-Clermont, M. (1991). "Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker for site-directed transformation of *Chlamydomonas." Nucl. Acids Res.* **19**(15): 4083-9.

Gorman, D. S. and R. P. Levine (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. (1992). "The endosymbiont hypothesis revisited." *Int. Rev. Cytol.* **141**: 233-237.

Grebenok, R. J., E. Pierson, et al. (1997). "Green fluorescent protein fusions for efficient characterization of nuclear targeting." *Plant J.* **11**(3): 573-86.

Gruber, H., S. H. Kirzenger, et al. (1996). "Expression of the *Volvox* gene encoding nitrate reductase: mutation-dependant activation of cryptic splice sites and intron-enhanced gene expression from a cDNA." *Plant Mol. Biol.* **31**: 1-12.

Gruenbaum, Y., T. Naveh-Many, et al. (1981). "Sequence specificity of methylation in higher plant DNA." *Nature* **292**: 860-862.

Gumpel, N. J. and S. Purton (1994). "Playing tag with *Chlamydomonas." Trends Cell Biol.* **4**: 299-301.

Gumpel, N. J., L. Ralley, et al. (1995). "Nuclear mutants of *Chlamydomonas reinhardtii* defective in the biogenesis of the cytochrome b6f complex." *Plant Mol. Biol.* 29(5): 921-32

Gumpel, N. J., J. D. Rochaix, et al. (1994). "Studies on homologous recombination in the green alga *Chlamydomonas reinhardtii.*" *Curr. Genet.* 26(5-6): 438-42.

Hall, L. M., K. B. Taylor, et al. (1993). "Expression of a foreign gene in *Chlamydomonas reinhardtii.*" *Gene* 124(1): 75-81.

Haring, M. A. and C. F. Beck (1997). "A promoter trap *Chlamydomonas reinhardtii*: development of a gene cloning method using 5' RACE based probes." *Plant J.* 11: 1341-1348.

Harris, E. H. (1984). "The *Chlamydomonas* Genetics Centre at Duke University." *Plant Mol. Biol. Rep.* 2: 29-41.

Harris, E. H. (1989). "The *Chlamydomonas* Sourcebook: a comprehensive guide to biology and laboratory use." San Diego, California, USA, Academic Press, Inc.

Haseloff, J. and B. Amos (1995). "GFP in plants." *Trends Genet.* 11(8): 328-329.

Haseloff, J., K. R. Siemering, et al. (1997). "Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly." *Proc. Natl. Acad. Sci. USA* 94(6): 2122-2127.

Hasnain, S. E., E. K. Manavathu, et al. (1985). "DNA-mediated transformation of *Chlamydomonas reinhardii* cells: use of aminoglycoside 3'-phosphotransferase as a selectable marker." *Mol. Cell. Biol.* 5(12): 3647-3650.

Hastings, J. W. (1996). "Chemistries and colors of bioluminescent reactions: a review." *Gene* 173: 5-11.

Heim, R., A. B. Cubitt, et al. (1995). "Improved green fluorescence." *Nature* 373: 663-664.

Heim, R., D. C. Prasher, et al. (1994). "Wavelength mutations and posttranslational autoxidation of green fluorescent protein." *Proc. Natl. Acad. Sci. USA* 91(26): 12501-12504.

Heinlein, M., B. L. Epel, et al. (1995). "Interaction of tobamovirus movement proteins with the plant cytoskeleton." *Science* 270: 1983-1985.

Hill, K. L., H. H. Li, et al. (1991). "Isolation and structural characterisation of the *Chlamydomonas reinhardtii* gene for cytochrome c6." *J. Biol. Chem.* 266: 15060-15067.

Ho, S. N., H. D. Hunt, et al. (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." *Gene* 77: 51-59.

Hodkinson, S. (1995). "GFP in *Dictyostelium.*" *Trends Genet.* 11(8): 327-328.

Hodson, R. C. and P. M. Gresshoff (1987). "Fluoro-acetamide resistance mutations in *Chlamydomonas reinhardtii.*" *Arch. Microbiol.* 148: 8-13.

Hofte, H. (1993). "An inventory of 1152 of expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana.*" *Plant J.* 4(6): 1051-1061.

Holliday, R. (1987). "The inheritance of epigenetic effects." *Science* **238**: 163-170.

Hoof, A. V. and P. J. Green (1996). "Premature nonsense codons decrease the stability of phytohemagglutinin mRNA in a position dependant manner." *Plant J.* **10**(3): 415-424.

Hoog, C. (1991). "Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy." *Nucl. Acids Res.* **19**(22): 6123-6127.

Hu, W. and C.-L. Cheng (1995). "Expression of *Aequorea* green fluorescent protein in plant cells." *FEBS Letts.* **369**: 331-334.

Hwang, S. and D. L. Herrin (1993). "Characterisation of a cDNA encoding the 20-kDa photosystem I light-harvesting polypepetide of *Chlamydomonas reinhardtii.*" *Curr. Genet.* **23**: 512-517.

Ikemura, T. (1981). "Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes." *J. Biol. Chem.* **146**: 1-21.

Ikemura, T. (1982). "Correlation between the abundance of yeast transfer RNAs and the occurence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs." *J. Mol. Biol.* **158**: 573-597.

Iman, S. H. and W. J. Snell (1988). "The *Chlamydomonas* cell wall degrading enzyme, lysin, acts on two substrates within the framework of the wall." *J. Cell Biol.* **106**: 2211-2221.

Imbault, P., C. Wittemer, et al. (1988). "Structure of the *Chlamydomonas reinhardtii cab-II-I* gene encoding a chlorophyll-a/b-binding protein." *Gene* **73**: 397-407.

Izant, J. G. and H. Weintraub (1984). "Inhibition of thymidine kinase gene expression by antisense RNA: a molecular approach to genetic analysis." *Cell* **36**: 1007-1015.

Jarvik, J. W., S. A. Adler, et al. (1996). "CD-tagging: a new approach to gene and protein discovery and analysis." *BioTechniques* **20**(5): 896-904.

Jarvis, E. E. and L. M. Brown (1991). "Transient expression of firefly luciferase in protoplasts of the green alga *Chlorella ellipsoidea.*" *Curr. Genet.* **19**: 317-321.

Jefferson, R. A. (1989). "The GUS reporter gene system." *Nature* **342**: 837-838.

Johnson, K. A. (1995). "Keeping the beat: form meets function in the *Chlamydomonas* flagellum." *BioEssays* **17**: 847-854.

Kaether, C. and H.-H. Gerdes (1995). "Visualisation of protein transport along the secretory pathway using green fluorescent protein." *FEBS Letts.* **369**: 267-271.

Khrebtukova, I. and R. J. Spreizer (1996). "Elimination of the *Chlamydomonas* gene family that encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxegenase." *Proc. Natl. Acad. Sci. USA* **93**: 13689-13693.

Kindle, K., K. L. Richards, et al. (1991). "Engineering of the chloroplast genome: Techniques and capabilites for chloroplast transformation in *Chlamydomonas reinhardtii.*" *Proc. Natl. Acad. Sci. USA* **88**: 1721-1725.

Kindle, K. L. (1990). "High-frequency nuclear transformation of *Chlamydomonas reinhardtii.*" *Proc. Natl. Acad. Sci. USA* **87**(3): 1228-32

Kindle, K. L., R. A. Schnell, et al. (1989). "Stable nuclear transformation of *Chlamydomonas* using the *Chlamydomonas* gene for nitrate reductase." *J. Cell. Biol.* **109**(6): 2589-2601.

Kindle, K. L. and O. A. Sodeinde (1994). "Nuclear and chloroplast transformantion in *Chlamydomonas reinhardtii*: strategies for genetic manipulation and gene expression." *J. Appl. Phycol.* **6**: 231-238.

Klaas, M., M. C. John, et al. (1989). "Rapid induction of genomic methylation and T-DNA gene expression in plant cells by 5-azacytosine derivatives." *Plant Mol. Biol.* **12**: 413-423.

Klein, T. M., E. D. Wolf, et al. (1987). "High-velocity microprojectiles for delivering nucleic acids into living cells." *Nature* **327**: 70-73.

Knaff, D. B. and M. Hirasawa (1991). "Ferredoxin-dependant chloroplast enzymes." *Biochim. Biophys. Acta* **1056**: 93-125.

Kobl, I., B. Scheidlmeier, et al. (1996). Poster: "Antisense inhibition of arylsulphatase in *Volvox* transformants." *Seventh International Conference on the Cell and Molecular Biology of Chlamydomonas.* University of Regensberg, Regensberg, Germany.

Koziel, M. G., N. B. Carozzi, et al. (1996). "Optimizing expression of transgenes with an emphasis on post-transcriptional events." *Plant Mol. Biol.* **32**: 393-405.

Kozminski, K. G., D. R. Diener, et al. (1993). "High level expression of nonacetylable α-tubulin in *Chlamydomonas reinhardtii.*" *Cell. Motil. Cytoskeleton* **25**: 158-170.

Krol, A. L. V. D., Mol J. N. M., et al. (1988). "Antisense genes in plants: an overview." *Gene* **72**: 45-50.

Lammertyn, E., L. V. Mellaert, et al. (1996). "Codon adjustment to maximise heterologous gene expression in *Streptomyces lividans* can lead to decreased mRNA stability and protein yield." *Mol. Gen. Genet.* **250**: 223-229.

Leeds, P., S. M. Peltz, et al. (1991). "The product of the yeast UPF1 gene is required for the rapid turnover of mRNAs containing a premature translational termination codon." *Genes Devel.* **5**: 2303-2314.

Levine, R. P. (1968). "Genetic dissection of photosynthesis." *Science* **162**: 768-771.

Lewski, B. K., R. Mah, et al. (1993). "Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3(2)." *J. Bacteriol.* **175**: 1995-2005.

Li, H. H., J. Quinn, et al. (1996). "Molecular genetic analysis of plastocyanin biosynthesis in *Chlamydomonas reinhardtii.*" *J. Biol. Chem.* **271**(49): 31283-31289.

Liss, M., D. L. Kirk, et al. (1997). "Intron sequences provide a tool for high-resolution phylogenetic analysis of volvicine algae." *Curr. Genet.* **31**: 214-227.

Luherson, K. R., S. Taha, et al. (1994). "Nuclear pre-mRNA processing in higher plants." *Progr. Nucl. Acids Res. Mol. Biol.* **47**: 149-193.

Lumbreras, V., D. R. Stevens, et al. (1998). "Efficient foreign gene expression in *Chlamydomonas reinhardtii* is mediated by an endogenous intron." *Plant J.* **14**:In press.

Mansour, S. L., K. R. Thomas, et al. (1988). "Disruption of the proto-oncogene int-2 in mouse embryo-derived stem-cells - a general strategy for targeting mutations to non-selectable genes." *Nature* **336**: 348-352.

Maquat, L. E. (1995). "When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells." *RNA* **1**: 453-465.

Mayfield, S. P. (1991). "Over-expression of the oxygen-evolving enhancer 1 protein and its consequences on photosystem II accumulation." *Planta* **185**: 105-110.

Mayfield, S. P., P. Bennoun, et al. (1987). "Expression of the nuclear encoded OEE1 protein is required for oxygen evolution and stability of photosystem II particles in *Chlamydomonas reinhardtii.*" *EMBO J.* **6**(2): 313-318.

Mayfield, S. P. and K. L. Kindle (1990). "Stable nuclear transformation of *Chlamydomonas reinhardtii* by using a *C. reinhardtii* gene as the selectable marker." *Proc. Natl. Acad. Sci. USA* **87**(6): 2087-2091.

Mayfield, S. P., M. Schirmer-Rahire, et al. (1989). "Analysis of the genes of the OEE1 and OEE3 proteins of the photosystem II complex of *Chlamydomonas reinhardtii.*" *Plant Mol. Biol.* **12**: 683-693.

McCombie, E. A. W. R. (1992). "*Caenorhabiditis elegans* expressed sequence tags identify gene families and potential disease gene homologues." *Nature Genet.* **1**: 124-131.

McHugh, J. P. and J. G. Spanier (1994). "Isolation of cadmium sensitive mutants in *Chlamydomonas reinhardtii* by transformation/insertional mutagenesis." *FEMS Microbiol. Letts.* **124**: 239-244.

McKinney, E. C., N. Ali, et al. (1995). "Sequence based identification of T-DNA insertion mutations in *Arabidopsis:* actin mutants act2-1 and act4-1." *Plant J.* **8**: 613-622.

Memon A. R., S. Hwang, et al. (1995). "Novel aspects of the regulation of a cDNA (Aff1) from *Chlamydomonas* with high sequence identity to animal ADP-ribosylation factor 1." *Plant Mol. Biol.* **29**: 567-577.

Michaelis, G., C. Varenholz, et al. (1990). "Mitochondrial DNA of *Chlamydomonas reinhardtii:* the gene for apocytochrome b and the complete function map of the 15.8 kb DNA." *Mol. Gen. Genet.* **223**: 211-216.

Millar, A. J., Carre I. A., et al. (1995). "The regulation of circadian period by phototransduction pathways in *Arabidopsis.*" *Science* **267**: 1161-1163.

Min'ko, I., S. Holloway, et al. (1998). Abstract: "Coral luciferase as a vital marker for chloroplast gene expression." *The VIIIth International Symposium on the Cell and Molecular Biology of Chlamydomonas.*, Lake Tahoe, California, 1988.

Mol, J. N. M., R. V. Blockland, et al. (1994). "Post-transcriptional inhibition of gene expression: sense and antisense genes." *Gene Inactivation and Homologous Recombination in Plants*. Ed: J. Paszowski. Dordrecht, Kluwer: 309-334.

Morrison, A., R. Marschalek, et al. (1997). "A novel, negative selectable marker for gene disruption in *Dictyostelium*." *Gene* **202**: 171-176.

Muller, F. W., G. L. Igloi, et al. (1992). "Structure of a gene encoding heat-shock protein HSP70 from the unicellular alga *Chlamydomonas reinhardtii*." *Gene* **111**(2): 165-73

Myster, S. H., J. A. Knott, et al. (1997). "The *Chlamydomonas* Dhc1 gene encodes a dynein heavy chain subunit required for assembly of the inner arm complex." *Mol. Cell. Biol.* **8**: 607-620.

Nasr, F., Betauche N., et al. (1994). "Heterospecific cloning of *Arabidopsis thaliana* cDNAs by direct complementation of pyrimidine auxotrophic mutants of *Saccharomyces cerevisiae*." *Mol. Gen. Genet.* **244**: 23-32.

Nelson, J. A. and P. A. Lefebvre (1995). "Targeted disruption of the NIT8 gene in *Chlamydomonas reinhardtii*." *Mol. Cell. Biol.* **15**(10): 5762-5769.

Nelson, J. A., P. B. Savereide, et al. (1994). "The CRY1 gene in *Chlamydomonas reinhardtii*: structure and use as a dominant selectable marker for nuclear transformation." *Mol. Cell. Biol.* **14**(6): 4011-4019.

Newman, S. M., Gillham N. W., et al. (1991). "Targeted disruption of chloroplast genes in *Chlamydomonas reinhardtii*." *Mol. Gen. Genet.* **230**(1-2): 65-74.

Niedz, R. P., Sussman M. R., et al. (1995). "Green fluorescent protein: an *in vivo* reporter of plant gene expression." *Plant Cell Rep.* **14**: 403-406.

O'Connor, H. E., S. V. Ruffle, et al. (1998). "The 9 kDa phosphoprotein of photosystem II. Generation and characterisation of *Chlamydomonas* mutants lacking PS-II-H and a site-directed mutant lacking the phosphorylation site." *Biolchim. Biophys. Acta.* In press.

Ohresser, M., R. F. Matagne, et al. (1997). "Expression of the arylsulphatase reporter gene under the control of the nit1 promoter in *Chlamydomonas reinhardtii*." *Curr. Genet.* **31**(3): 264-271.

Oparka, K. J., A. G. Roberts, et al. (1995). "Imaging the green fluorescent protein in plants - viruses carry the torch." *Protoplasma* **189**: 133-141.

Ormo, M., A. B. Cubitt, et al. (1996). "Crystal structure of the *Aequorea victoria* green fluorescent protein." *Science* **273**(5280): 1392-5.

Ow, D. W., K. V. Wood, et al. (1986). "Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants." *Science* **234**: 856-859.

Palomares, R., R. G. Herrmann, et al. (1993). "Antisense RNA for components associated with the oxygen-evolving complex and the Rieske iron/sulfur protein of the tabacco thylakoid membrane suppresses the accumulation of mRNA, but not of protein." *Planta* **190**: 305-312.

Pang, S. Z., D. L. DeBoer, et al. (1996). "An improved green fluorescent protein as a vital marker in plants." *Plant Physiol.* **112**(3): 893-900.

Parzour, G. J., O. A. Sineshchekov, et al. (1995). "Mutational analysis of the phototransduction pathway of *Chlamydomonas reinhardtii.*" *J. Cell Biol.* **131**: 427-440.

Pascher, A. (1918). "Uber die Beziehung der Reduktionsteilung zur Mendelschen Spatulung." *Ber. Dtsch. Bot. Ges.* **36**: 163-168.

Pearson, W. R. and D. J. Lipman (1988). "Improved tools for biological sequence comparison." *Proc. Natl. Acad. Sci. USA* **85**: 2444-2448.

Perlak, F. J., R. L. Fuchs, et al. (1991). "Modification of the coding sequence enhances plant expression of insect protein control genes." *Proc. Natl. Acad. Sci. USA* **88**: 3324-3328.

Pines, J. (1995). "GFP in mammalian cells." *Trends Genet.* **11**(8): 326-327.

Prasher, D. C. (1995). "Using GFP to see the light." *Trends Genet.* **11**(8): 320-323.

Prasher, D. C., V. K. Eckenrode, et al. (1992). "Primary structure of the *Aequorea victoria* green fluorescent protein." *Gene* **111**: 229-33.

Prieto, R., J. M. Pardo, et al. (1996). "Salt sensitive mutants of *Chlamydomonas reinhardtii* isolated after insertional tagging." *Plant Physiol.* **112**: 99-104.

Pulak, R. and P. Anderson (1993). "mRNA surveilance by the *Caenorhabditis elegans* smg genes." *Genes Devel.* **7**: 1885-1897.

Purton, S. (1995). "The chloroplast genome of *Chlamydomonas.*" *Sci. Prog.* **78**(3): 205-216.

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

Purton, S. and J. D. Rochaix (1994). "Complementation of a *Chlamydomonas reinhardtii* mutant using a genomic cosmid library." *Plant Mol. Biol.* **24**(3): 533-537.

Radford, A. and N. I. M. Dix (1988). "Comparison of the orotidine-5'-monophosphate decarboxylase sequences of eight species." *Genome* **30**: 501-505.

Radich, J. (1996). "Commercial implications of the Human Genome Project." *TIBS/TECH* **14**: 143-147.

Randolph-Anderson, B. L., J. E. Boynton, et al. (1993). "Futher characterisation of the respiratory deficient *dum-1* mutation of *Chlamydomonas reinhardtii* and its use as a recipient for mitochondrial transformation." *Mol. Gen. Genet.* **236**: 235-244.

Ranum, L. P. W., M. D. Thompson, et al. (1988). "Mapping flagellar genes in *Chlamydomonas* using Restriction Length Fragment Polymorphisms." *Genetics* **120**: 109-122.

Reinhardt, L. (1876). "Die Kopulation bei *Chlamydomonas pulvisculus* Ehr. und *Stigeoclonium* sp." *Arb. Naturforsch. Ges. Univ. Charkow* **10**.

Renckens, S., H. D. Greve, et al. (1992). "Petunia plants escape from negative selection against a transgene by silencing the foreign DNA via methylation." *Mol. Gen. Genet.* **233**: 53-64.

Rizzuto, R., M. Brini, et al. (1995). "Chimeric green fluorescent protein as a tool for visualizing subcellular organelles in living cells." *Curr. Biol.* **5**(6): 635-642.

Robinson, C. and A. Mant (1997). "Targeting of proteins into and across the thylakoid membrane." *Trends Plant Sci.* **2**(11): 431-437.

Robinson, M., R. Lilley, et al. (1984). "Codon usage can affect efficiency of translation ofgenes in *E. coli.*" *Nucl. Acids Res.* **12**: 6663-6671.

Rochaix, J.-D. and J. V. Dillewijn (1982). "Transformation of the green alga *Chlamydomonas reinhardtii* with yeast DNA." *Nature* **290**: 70-72.

Rochaix, J.-D., S. Mayfield, et al. (1988). 'Molecular biology of *Chlamydomonas.*" *Plant Molecular Biology, A Practical Approach.* Ed: C. H. Shaw. Oxford, UK, IRL Press: 253-275.

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

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

Rouwendal, G. J., O. Mendes, et al. (1997). "Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage." *Plant Mol. Biol.* **33**(6): 989-999.

Sagar, R. and G. E. Palade (1967). "Structure and development of the chloroplast in *Chlamydomonas.* The normal green cell." *J. Biochem. Cytol.* **3**: 463-487.

Sakamoto, W., K. L. Kindle, et al. (1993). "*In vivo* analysis of *Chlamydomonas* chloroplast *petD* gene expression using stable transformation of beta-glucuronidase translational fusions." *Proc. Natl. Acad. Sci. USA* **90**(2): 497-501

Sakamoto, W., N. R. Sturm, et al. (1994). "PetD mRNA maturation in Chlamydomonas reinhardtii chloroplasts: role of 5' endonucleolytic processing." *Mol. Cell. Biol.* **41**(9): 6180-6186.

Sambrook, J., E. F. Fritsch, et al. (1989). *Molecular Cloning: A Laboratory Manual,* 2nd edition. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." *Proc. Natl. Acad. Sci. USA* **74**(12): 5463-5467.

Savard, F., C. Richard, et al. (1996). "The *Chlamydomonas reinhardtii L1818* gene represents a distant relative of the *cabI/II* genes that is regulated during the cell cycle and in response to illumination." *Plant Mol. Biol.* **32**: 461-473.

Schägger, H. and G. von Jagow (1987). "Tricine-sodium dodecyl sulphate polyacrylamide gel electrophoresis for the separation of proteins in the range from 1-100kDa." *Anal. Biochem.* **166**: 368-379.

Schena, M. (1996). "Genome analysis with gene-expression microarrays." *BioEssays* **18**(5): 427-431.

Schena, M., D. Shalon, et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary cDNA microarray." *Science* **270**: 467-470.

Schnell, R. A. and P. A. Lefebvre (1993). "Isolation of the *Chlamydomonas* regulatory gene NIT2 by transposon tagging." *Genetics* **134**: 737-747.

Sheen, J., S. Hwang, et al. (1995). "Green fluorescent protein as a vital marker in plant cells." *Plant J.* **8**: 777-784.

Shimomura, O., F. H. Johnson, et al. (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan *Aequorea.*" *J. Cell Comp. Physiol.* **59**: 223-239.

Silflow, C. D., P. Kathir, et al. (1995). "Molecular mapping of genes for flagellar proteins in *Chlamydomonas.*" *Methods Cell Biol.* **4**: 525-530.

Simons, R. W. (1988). "Naturally occuring antisense RNA control-a brief review." *Gene* **72**: 35-44.

Sizova, E. A. I. A. (1996). "Stable nuclear transformation of *Chlamydomonas reinhardtii* with a *Streptomyces rimosus* gene as the selective marker." *Gene* **181**: 13-18.

Sodeinde, O. A. and K. L. Kindle (1993). "Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii.*" *Proc. Natl. Acad. Sci. USA* **90**(19): 9199-9203.

Spurr, N. K. (1995). "Navigating the genome." *Biochem. Soc. Trans.* **24**: 285-289.

Stearns, T. (1995). "Green fluorescent protein. The green revolution." *Curr. Biol.* **5**(3): 262-4.

Stemmer, W. P. C. (1994). "DNA shuffling by random fragmentation and assembly: *in vitro* recombination for molecular evolution." *Proc. Natl. Acad. Sci. USA* **91**: 10747-10751.

Stevens, D. R., S. Purton, et al. (1996). "The bacterial phleomycin resistance gene *ble* as a dominant selectable marker in *Chlamydomonas.*" *Mol. Gen. Genet.* **251**(1): 23-30.

Strachan, T., M. Abitol, et al. (1997). "A new dimension for the human genome project: toward comprehensive expression maps." *Nature Genet.* **16**: 1267-1320.

Sullivan, M. L. and P. J. Green (1993). "Post-transcriptional regulation of nuclear-encoded genes in higher plants: the roles of mRNA stability and translation."*Plant Mol. Biol.* **23**: 1091-1104.

Tam, L. W. and P. A. Lefebvre (1993). "Cloning of flagellar genes in *Chlamydomonas reinhardtii* by DNA insertional mutagenesis." *Genetics* **135**(2): 375-384.

Tam, L. W. and P. A. Lefebvre (1995). "Insertional mutagenesis and isolation of tagged genes in *Chlamydomonas.*" *Methods Cell. Biol.* **47**: 519-523.

Tang, D. K., S. Y. Qiao, et al. (1995). "Insertional mutagenesis of *Chlamydomonas reinhardtii* by electroporation and heterologous DNA." *Biochem. Mol. Biol. Int.* **36**(5): 1025-1035.

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

Turner, M., N. Gumpel, et al. (1996). "Biogenesis of the electron-transfer complexes in *Chlamydomonas reinhardtii.*" *Biochem. Soc. Trans.* **24**(3): 733-8.

Uchimiya (1992). "Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice." *Plant J.* **2**: 1005-1009.

Ueda, Y., S. Taguchi, et al. (1993). "Effect of a rare leucine codon, TTA, on expression of a foreign gene in *Streptomyces lividans.*" *Biol. Biophys. Acta* **1172**: 262-266.

Vashishtha, M., G. Segil, et al. (1996). "Direct complementation of *Chlamydomonas* mutants with amplified YAC DNA." *Genomics* **36**(3): 459-67.

Walbot, V. (1992). "Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis." *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**: 49-82.

Wilkerson, C. G., S. M. King, et al. (1995). "The 78000 m_r intermediate chain of *Chlamydomonas* outer arm dynein is a WD-repeat protein required for arm assembly." *J. Cell Biol* **129**(1): 169-178.

Witman, G. B. (1993). "*Chlamydomonas* phototaxis." *Trends Cell Biol.* **3**: 403-408.

Wolfsberg, T. G. and D. Landsman (1997). "A comparison of expressed sequence tags (ESTs) to human genomic sequences." *Nucl. Acids Res.* **25**(8): 1626-1632.

Woolf, T. (1996). "It's not the size, it's the potency." *Nature Biotech.* **14**: 824-844.

Worley, K. C., B. A. Wiese, et al. (1995). "BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results." *Genome Res.* **5**: 173-184.

Yang, T. T., S. R. Kain, et al. (1996). "Dual color microscopic imagery of cells expressing the green fluorescent protein and a red-shifted variant." *Gene* **173**: 19-23.

Zhang, G., V. Gurtu, et al. (1996). "An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells." *Biochem. Biophys. Res. Commun.***227**(3): 707-711.

Zhang, H., P. L. Herman, et al. (1994). "Gene isolation through genomic complementation using an indexed library of *Chlamydomonas reinhardtii* DNA." *Plant Mol. Biol.* **24**: 663-672.

Zhang, D., P. A. Lefebvre, (1997). "FAR1, a negative regulatory locus required for the expression of the nitrase reductase gene in *Chlamydomonas reinhardtii.*" *Genetics* **146**: 121-133.

Zolotukhin, S., M. Potter, et al. (1996). "A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells." *J. Virol.* **70**(7): 4646-54.