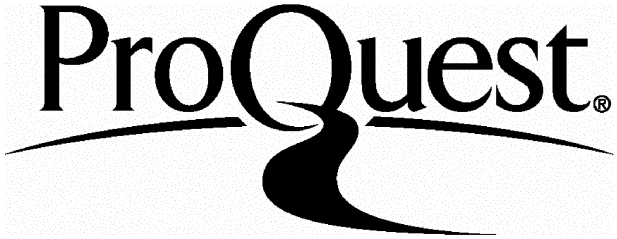
A Functional Study of Bacterial Orthologues of the Malarial Plastid Gene, *ycf* 24

A thesis submitted for the degree of
Doctor of Philosophy
by

Anna Elizabeth Law

Division of Parasitology
National Institute for Medical Research



Abstract

Malaria parasites have two DNA containing organelles, a mitochondrion and a plastid. Most of the plastid genes are involved in gene expression however there are a number of open reading frames (ORFs) of unknown function. In *Plasmodium falciparum*, ORF470 is particularly interesting as it is highly conserved with an orthologue (*ycf* 24) in both algae and bacteria. Some annotations describe *ycf* 24's product as an ABC transporter subunit but the level of significance is low.

To investigate the function of *ycf* 24, the cyanobacterium *Synechocystis*, strain PCC6803, was transformed with a disrupted version of the gene and allowed to undergo several rounds of replication before selection with kanamycin. *Synechocystis* 6803 was chosen as it carries up to 10 copies of its genome per cell, allowing the support and selection of lethal genes. Primary transformants formed "ragged" colonies attributable to the death of homoplasmic transformants (complete "knockouts"). This indicates that loss of the gene is lethal. After selection, transformants were shown to be heteroplasmic (containing both disrupted and wild type *ycf* 24). Furthermore, heteroplasmic cells plated in the absence of kanamycin resumed WT growth rate and colony morphology indicating that the presence of the disrupted gene was deleterious and/or that the gene product is required in stoichiometric amounts.

When the gross morphology of cells from ragged colonies was observed by scanning electron microscopy, a statistically significant accumulation of cells in the last stages of cell division was apparent. This suggests that disruption of *ycf* 24 leads to a defect in cytokinesis. Upon relaxation of kanamycin selection, the proportion of dividing cells returned to one resembling WT.

Attempts were made to knock out the single copy of *ycf* 24 in *E. coli*. One approach made use of the temperature sensitive gene replacement vector pKO3 that contains genes for chloramphenicol resistance and levansucrase, which is lethal to *E. coli* in the presence of sucrose. This protocol allowed both positive and negative selection of recombination at the *ycf* 24 locus. However, no viable knockouts were obtained.

Over-expression of *ycf* 24 in *E. coli* using the pMAL-c2 vector resulted in moderate filamentation, the fusion protein localizing as a band on either side of the nucleoid, sometimes extending along the membrane to the cell poles. Unlike some

other partition mutants, FtsZ ring formation was inhibited in such cells and the first division of nucleoids following induction resulted in their accumulation transverse to the cell's length rather than longitudinally in the usual way.

It is proposed that *ycf*24 is an essential prokaryotic gene involved in cell division and/or DNA segregation.

Acknowledgements

Primarily I would like to thank my supervisor Iain Wilson for providing me with support and direction throughout my course of study. I would also like to thank Conrad Mullineaux at University College London for introducing me to cyanobacteria, and for all his help. For their invaluable advice, instruction and friendship I want to thank all the other members of the lab at NIMR; Barbara Clough, Katherine Ellis, Kaveri Rangachari, Peter Preiser and Don Williamson and also Malcolm Strath for his help with protein purification, for running sequencing gels and for antibody production. In addition, I thank Liz Hirst at NIMR for the electron microscopy work and acknowledge Nick Fisher (UCL) and Elaine Davis (NIMR) for their advice on my *E. coli* gene knockout studies.

For donating the antibiotic resistance genes, as well as his help, I would like to thank Mark Ashby (UCL). Likewise I'd like to thank Daniel Emlyn-Jones (UCL) for the kanamycin resistant *Synechocystis* strain; Elaine Davis (NIMR) for the KM4104 *E. coli* strain; and George Church (Harvard University) for the pKO3 vector

On a more personal note I owe a lot to Pete - I could not have done this without him. Finally, I thank my parents for their support and encouragement through thick and thin.

Contents

List of Figures

List of Tables

Abbreviations

A	Absorbance at wavelength (nm) stated
AA~tRNA	Aminoacyl tRNA
Amp	Ampicillin
ACP	Acyl carrier protein
ATP	Adenosine 5′-triphosphate
BG11	Blue green 11 medium
bp	Base pairs
BSA	Bovine serum albumen
Cm $^{(r \text{ or } s)}$	Chloramphenicol (resistance or sensitivity)
CTAB	Cetyl trimethyl ammonium bromide
DAPI	4′,6-diamidinio-2-phenylidole dihydrochloride hydrate
dATP	2′-deoxyadenosine-5′-triphosphate
DEAE	Diethylaminoethyl
dH_2O	Distilled water
DNA	Deoxyribose nucleic acid
DNase	Deoxyribonuclease
EDTA	Ethylenediaminetetra-acetic acid
EF-G	Elongation factor G
EF-Tu	Elongation factor Tu
ER	Endoplasmic reticulum
GST	Glutathione-S-transferase
HMDS	Hexamethyl disilazane
HRP	Horse radish peroxidase
Ig	Immunoglobulin
IPTG	Isopropylthio-β-D-galactoside
IR $_{(A \text{ or } B)}$	Inverted repeat (1^{st} or 2^{nd})
kb	Kilobase(s)
kDa	Kilodaltons
Km$^{(r \text{ or } s)}$	Kanamycin (resistance or sensitivity)
LB agar	Luria-Bertani agar
LB broth	Luria-Bertani medium
LSU	Large subunit

M	Marker
MBP	Maltose binding protein
MOPS	3-[N-Morpholono]propane sulphonic acid
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NIMR	National Institute for Medical Research
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PD	Plastid dividing
plDNA	Plastid DNA
PS	Photosystem
RNA	Ribose nucleic acid
RNase	Ribonuclease
rp (l or s)	Ribosomal protein genes (large or small)
rpm	Revolutions per minute
rRNA	Ribosomal ribose nucleic acid
RT-PCR	Reverse transcriptase PCR
SDS	Sodium dodecyl sulphate
SE	South East
SEM	Scanning electron microscope (microscopy)
Sp1	Streptomycin/spectinomycin resistance (*aadA* gene) cassette
SSC	Saline sodium citrate
SSPE	Sodium chloride sodium phosphate EDTA
SSU	Small subunit rRNA
Strep^r	Streptomycin (resistance)
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TE	Tris EDTA
TEM	Transmission electron microscope (microscopy).
TES	N-tris(hydroxymethy)methyl-2-aminoethanesulfonic acid
Tris	Tris(hydroxymethyl)methylamine

tRNA	Transfer ribonucleic acid
U	Unit(s)
UCL	University College London
UV	Ultraviolet
v/v	Volume by volume
WT	Wild type
w/v	Weight by volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
ycf	Hypothetical chloroplast open reading frame

Chapter 1. Introduction

1.1 Malaria

1.1.1 Protozoans and the phylum Apicomplexa

All members of the phylum Apicomplexa are obligate intracellular parasites with similar sexual and asexual stages in their life cycles (Levine, 1987). However, the groups within Apicomplexa are diverse and the phylogenetic relationships unclear. The only common feature is a set of specialized organelles (the apical complex) required for attachment to and penetration of host cells. Contained within the apical complex are polar ring(s), conoid micronemes, rhoptries and subpellicular microtubules. Some apicomplexans are important animal and human pathogens, including the AIDS-related opportunistic pathogens *Toxoplasma* and *Cryptosporidium*, and the economically important pathogens, *Theileria* and *Eimeria* of cattle and chicken, respectively. However, perhaps the most important of them all is the genus *Plasmodium*, the causative agent of malaria.

1.1.2 *Plasmodium* sp. and the disease

Malaria remains a major disease with between 1.5 and 2.7 million fatalities per year, most of them young African children. As mentioned above, it is caused by a parasitic protozoan of the phylum Apicomplexa and is transmitted by the bite of the female *Anopheles* mosquito. The causative agent, *Plasmodium* sp. has a complex life cycle, consisting of alternating extra and intracellular stages in both the human host and mosquito vector (Fig. 1.1). Four species cause malaria in humans, *P. vivax*, *P. ovale*, *P. malariae* and *P. falciparum*, the main diagnostic feature being recurring bouts of fever the frequency depending upon the length of the erythrocytic cycle (Phase 4 in Fig. 1.1). However malaria can mimic the symptoms of many other diseases, eg. anaemia, jaundice, gastro-enteritis or influenza. In the case of *P. vivax* infection, a dormant form, hypnozoites, can lie in liver cells for months or years before causing the illness to relapse (Phase 3 (5) in Fig. 1.1). The most serious disease is attributed to *P. falciparum*, the causative agent of cerebral malaria. In *P. falciparum* infections the initial mild symptoms of malaria are suddenly replaced by signs of overwhelming illness including liver, kidney and respiratory failure or coma (Knell, 1991). The severity of disease caused by this species of *Plasmodium* results primarily from its ability to modify the surface of infected red blood cells by inserting parasite

Figure 1.1

The life cycle of *Plasmodium* (Knell, 1991).

Phase 1: Fertilization in the mosquito.

Following a blood meal from an infected person by a female *Anopheles* mosquito, gametocytes escape from the red blood cells. Fertilization occurs when male and female gametes meet, and the zygote develops into an invasive ookinete. This bores into the mosquito stomach wall and becomes an oocyst.

Phase 2: Sporogony: asexual development in the parasite.

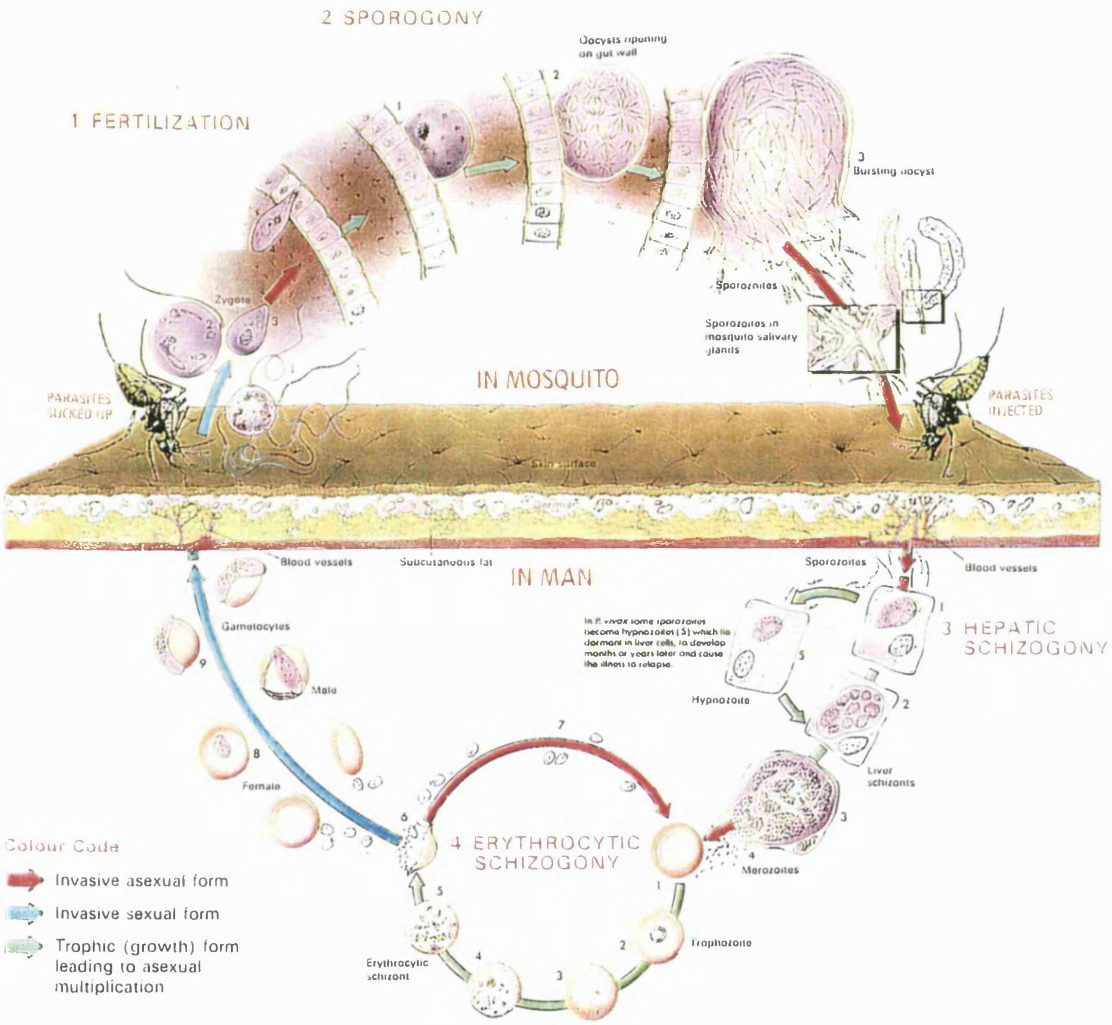
Following a phase of growth the oocyst divides into thousands of invasive sporozoites. The oocyst then bursts and the sporozoites migrate through the body of the mosquito and invade its salivary glands.

Phase 3: Hepatic schizogony; in the vertebrate host.

When an infected mosquito feeds, sporozoites enter the bloodstream and invade liver cells. These trophozoites grow and divide into thousands of invasive merozoites. Infected liver cells then burst and release merozoites into the blood. In *P. vivax* some sporozoites become hypnozoites which lie dormant in liver cells, to develop months or years later and cause the illness to relapse.

Phase 4: Erythrocytic schizogony; asexual development in the blood.

Merozoites invade red blood cells and become erythrocytic trophozoites. These grow, then divide into 8-16 new merozoites. Parasite maturation causes the red blood cell to burst and merozoites are released and the cycle starts again. As the disease progresses some merozoites develop into male or female gametocytes. These circulate in the blood, but only develop further if taken up by a mosquito.


2 SPOROGONY
Oocysts ripening on gut wall
1 FERTILIZATION
3 Bursting oocyst
Zygote
Sporozoites
Sporozoites in mosquito salivary glands
IN MOSQUITO
PARASITES SUCKED UP
PARASITES INJECTED
Skin surface
Blood vessels
Subcutaneous fat
IN MAN
Sporozoites
Blood vessels
Gametocytes
In P. vivax some sporozoites become hypnozoites (5) which lie dormant in liver cells, to develop months or years later and cause the illness to relapse.
3 HEPATIC SCHIZOGONY
Male
Hypnozoite
Female
Liver schizonts
4 ERYTHROCYTIC SCHIZOGONY
Merozoites
Trophozoite
Erythrocytic schizont
Colour Code
Invasive asexual form
Invasive sexual form
Trophic (growth) form leading to asexual multiplication

proteins. *P. falciparum* infected erythrocytes can bind to host endothelial cells, a process called cytoadherance, leading (in some cases) to the accumulation of infected red blood cells in specific organs. Cerebral malaria is caused by the accumulation of erythrocytes in the small blood vessels of the brain (MacPherson *et al.*, 1985). As a consequence, nearly all deaths from malaria are attributed to *P. falciparum* infection.

1.1.3 Treatment and eradication of malaria

Until recently the damage caused by malaria has been limited by the use of anti-malarial drugs such as chloroquine. However, resistance to this and other drugs is spreading, particularly in SE Asia and Africa (Miller, 1997). As a consequence new drugs are urgently required to replace those becoming obsolete. Control of the disease by targeting the mosquito vector, using methods such as insecticide-impregnated bed nets, is fraught with problems and often is hindered by social, economical and environmental factors (Knell, 1991). Another option is control by vaccination, and the potential success of vaccine development was indicated over 20 years ago when it was shown that individuals continually exposed to infection develop immunity. Cohen *et al.* (1961) showed that infants born in hyperendemic regions were "relatively resistant" within the first three months of life due to passive transfer of antibodies from the mother. Thereafter children suffered severe, recurrent attacks of the disease that became less frequent in later childhood and rare (in acute form) in adults (Wilson *et al.*, 1950; McGregor, 1960). In addition to naturally acquired immunity, inoculation of live attenuated parasites also offers protection (Clyde *et al.*, 1973) and protection by immunization with whole, killed organisms has had some success in animal models (Mitchell *et al.*, 1975). These positive observations led to the identification of many different antigens as candidates for subunit vaccines (Berzins and Perlmann, 1996) and some success has been observed in vaccination studies with single proteins or peptides. One example is the double polypeptide vaccine RTS,S produced by Stoute *et al.* (1998). However, although an immune response was observed, protection was short lived and a safe, effective vaccine has yet to be developed. This is not surprising due to the complexities of the life cycle of the parasite; the stages in humans for example are both morphologically and antigenically distinct with stage specific immunity. Because of this it seems likely that multistage, multivalent vaccines will need to be developed (Holder, 1999).

1.1.4 Malaria genome sequencing project

The genome of *P. falciparum* consists of 14 chromosomes, comprising ~5,000 genes which are extremely rich in adenine and thymine nucleotides (81-82%) (Hyde *et al.*, 1989). The entire genome is currently being sequenced by the Malaria Genome Sequencing Consortium. Collaborators include the Sanger Centre (UK), The Institute for Genomic Research, Naval Medical Research Institute (USA) and Stanford University (USA) (Wirth, 1998). Chromosome 2 has been completely sequenced (Gardner *et al.*, 1998), as has chromosome 3 (Bowman *et al.*, 1999), and it is hoped that the entire genome will be completed by the end of 2001 (personal communication from D. Lawson). With the completion of sequencing, the full spectrum of proteins expressed at each stage of the life cycle of the parasite can be determined. This should allow the development of more effective, stage-specific drugs or vaccines against malaria based on new parasite protein targets.

1.2 The 35kb extrachromosomal DNA element of malaria

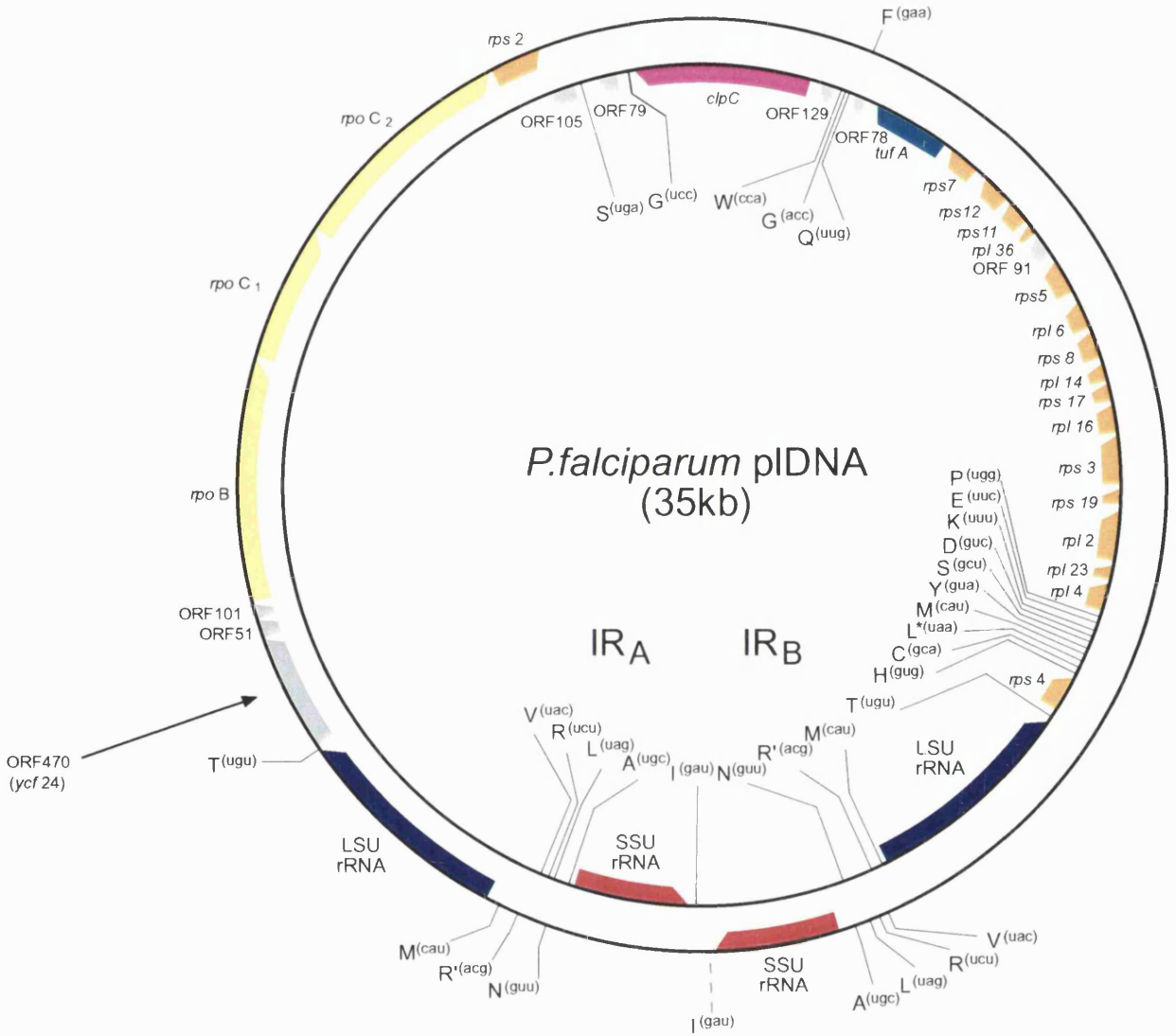
Malaria, along with several other apicomplexan parasites carries two extrachromosomal DNAs whose sizes are 6Kb and 35Kb respectively (Wilson *et al.*, 1991). The former corresponds to mitochondrial DNA whereas the latter is carried by a vestigial plastid organelle, and resembles the remnant of an algal plastid genome (Wilson *et al.*, 1994).

The mapping and complete sequencing of the malarial 35Kb DNA genome (Wilson *et al.*, 1996) revealed two main features. The first was an inverted repeat (IR) comprising 1/3 of the genome and the second, a remarkably high A + T content (86.9%). The complete gene map of the circular 35kb genome of *P. falciparum* is given in Figure 1.2. Many genes required for protein synthesis were identified on the plastid genome. These include genes for transfer RNA, 16S (SSU) and 23S (LSU) rRNAs and, small and large subunit ribosomal proteins. The circle also codes for three subunits of a prokaryotic-like RNA polymerase, the translation elongation factor Tu (coded for by *tuf* in *Escherichia coli*), and a probable molecular chaperone encoded by a version of the *clpC* gene. Lastly, eight unidentified potential open reading frames (ORFs) were identified on the circle.

Figure 1.2

Complete gene map of the circular 35kb genome from *Plasmodium falciparum* (Wilson *et al.*, 1996).

The two sectors (A and B) of the inverted repeat are indicated and tRNA genes are signified by both the single letter amino acid code and anticodon. The direction of transcription is anticlockwise (inner circle) and clockwise (outer circle). Unidentified open reading frames (ORFs) are numbered according to the length of the predicted peptide. The largest of these (shown by the arrow) is ORF 470, *ycf* 24, which has orthologues in a variety of bacteria. *rpl* and *rps* encode large subunit (LSU) and small subunit (SSU) ribosomal proteins respectively. *rpo* genes encode subunits of a "eubacteria-like" RNA polymerase; *tufA* encodes a translation elongation factor and *clpC*, a putative molecular chaperone.


P.falciparum plDNA
(35kb)
IR_A
IR_B
rps 2
rpo C 2
rpo C 1
rpo B
ORF101
ORF51
ORF470
(ycf 24)
T(ugu)
LSU rRNA
SSU rRNA
ORF105
ORF79
clpC
ORF129
ORF78
tuf A
F(gaa)
S(uga)
G(ucc)
W(cca)
G(acc)
Q(uug)
rps7
rps12
rps11
rpl 36
ORF 91
rps5
rpl 6
rps 8
rpl 14
rps 17
rpl 16
rps 3
rps 19
rpl 2
rpl 23
rpl 4
rps 4
P(ugg)
E(uuc)
K(uuu)
D(guc)
S(gcu)
Y(gua)
M(cau)
L*(uaa)
C(gca)
H(gug)
T(ugu)
V(uac)
R(ucu)
L(uag)
A(ugc)
I(gau)
N(guu)
R'(acg)
M(cau)
LSU rRNA
SSU rRNA
M(cau)
R'(acg)
N(guu)
I(gau)
SSU rRNA
A(ugc)
L(uag)
R(ucu)
V(uac)

1.2.1 Discovery

The first observation of the 35kb DNA element was probably in the early 1970s when, following preparation of a CsCl gradient of DNA from *P. knowlesi*, a low density molecule was observed (Gutteridge *et al.*, 1971). This was assumed to have arisen from the "mitochondrial-like" organelles seen in *Plasmodium coatneyi* from rhesus monkeys (Rudzinska and Trager, 1968) and from rodent malarial parasites (Howells and Fullard, 1970).

In 1975, electron micrographs of mitochondria-enriched extracts isolated from the avian malarial parasite *P. lophurae,* showed circular molecules of DNA (Kilejian, 1975) of ~10.3μm in contour length. Similar circular DNAs were later observed in another apicomplexan parasite, *Toxoplasma gondii* (Borst *et al.*, 1984). The molecules (13μm contour length) were found to have a large cruciform structure. A year later, two AT-rich DNA elements were isolated from the simian malarial parasite, *P. knowlesi*, (Williamson *et al.*, 1985) in a "light" DNA band corresponding to approximately 1% of the total DNA isolated. Like those described by Borst *et al.* (1984), the circular DNA molecules (of 11.6μm contour length) also had a cruciform structure. However, it was not until three years later that the extrachromosomal DNA elements were isolated from *P. falciparum* (Gardner *et al.*, 1988). Analysis showed one of these was a 6kb linear, tandemly repeated DNA (Vaidya *et al.*, 1989) and the other a 35kb circular DNA (Gardner *et al.*, 1988). Furthermore, restriction analysis and Southern hybridization of the latter indicated homology to the circular DNA of *P. knowlesi*. Figure 1.3 shows a molecule of 35kb circular DNA, with its incipient cruciform, isolated from *P. falciparum* (Wilson and Williamson, 1997).

1.2.2 Genetic evidence for a plastid origin

1.2.2.1 Plastid or mitochondrial?

Initially it was assumed that both of the malarial extra-chromosomal elements were mitochondrial. It was not until 1991 that Wilson *et al.* suggested that the 35kb DNA was plastid in origin. Original assumptions of a mitochondrial origin rested on the prokaryotic nature of the genes encoded by both the 6kb and 35kb elements (Gardner *et al.*, 1988; Vaidya, 1989). Moreover, both satellite DNAs contained a sequence with extensive homology to the small subunit ribosomal RNA of mitochondria. Much lower homology was observed with eukaryotic nuclear DNA

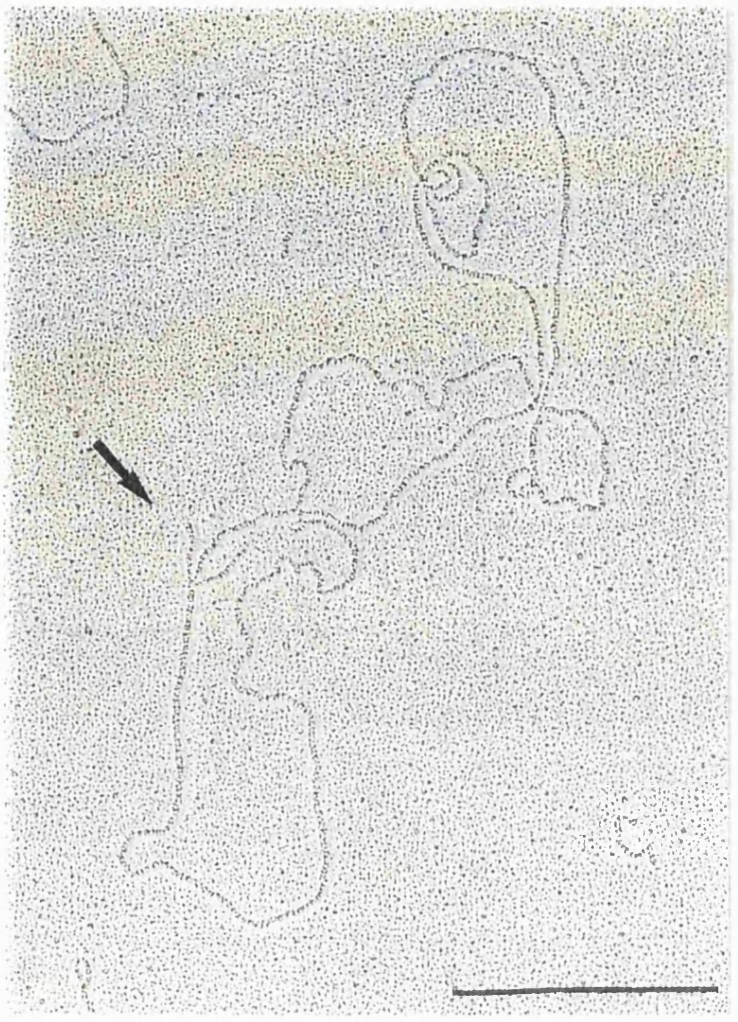


Figure 1.3

Electron micrograph of the 35kb circular extrachromosomal DNA molecule from *Plasmodium falciparum* (Wilson and Williamson, 1997).

A cruciform structure corresponding to the inverted rRNA repeat is shown by the arrow. The bar corresponds to 1μm.

(Gardner *et al.*, 1988). Bipartite mitochondrial genomes, although rare, are not unknown and have been observed in plants (Rush and Misra, 1985; Small *et al.*, 1989). They are thought to be the result of evolutionary genome reorganization, via recombination events between substoichiometric mitochondrial intermediates. These intermediates finally result in new mitochondrial genomes containing novel repeats or deletions. The large size and presence of large direct repeats which are common in the mitochondrial genomes of higher plants are assumed to be a consequence of this mode of reorganization (Palmer and Shields, 1984; Lonsdale *et al.*,1984). But the concept of mitochondrial intermediates did not fit the case for apicomplexan parasites as only the 6kb linear DNA co-fractionated with mitochondia (Wilson *et al.*, 1992). Moreover, a C residue, which occurs at C1160 (equivalent to position 1207 in *E. coli* 16S rRNA) was found in the SSU rRNA. This residue is found in all plastid SSU rRNAs but in none of the mitochondrial ones (Gardner *et al.*, 1991a). Finally, Feagin *et al.* (1992) showed that although both the 6kb and 35kb elements contained genes for "prokaryotic like" large and small subunit rRNAs, there was less conservation between them than to *E. coli* sequences. This suggested the two extrachromosomal DNA elements of *P. falciparum* were not closely related.

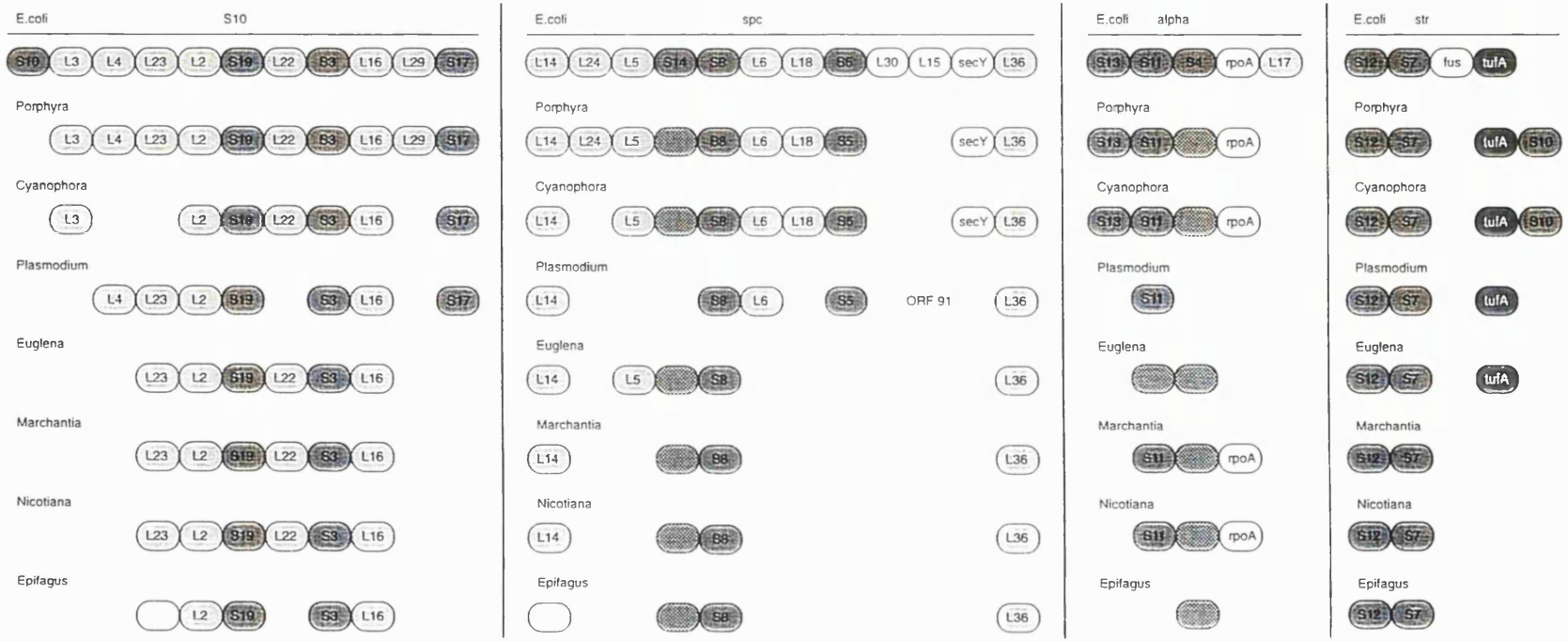
Another striking feature of the 35kb genome was the inverted repeat, which is typical of chloroplast genomes (Palmer, 1985) but never observed in mitochondrial genomes. The inverted rRNA repeat of *P. falciparum* is flanked by a *rpoBC* operon (subunits for RNA polymerase) on one side and clustered ribosomal genes on the other (Fig. 1.2), an organization typical of plastids (Gardner *et al.*, 1991b; Howe, 1992). Furthermore, the order of the ribosomal protein genes (encoding both the large and small subunits, *rpl* and *rps*) suggests a fusion of the *S10*, *spc*, *alpha* and *str* operons of *E. coli*, as found in other plastids (Figure 1.4; Wilson *et al.*, 1996).

The presence of the *rpo* genes, *rpoB* and *rpoC*, also was indicative of a plastid ancestry as they are absent from mitochondrial genomes, with the exception of the primative freshwater protozoan *Relictomonas americana* whose mitochondrial genome contains four genes specifying a multi-subunit eubacterial-type RNA polymerase (Lang *et al.*, 1997). All other mitochondria are thought to use a nuclear encoded single-subunit enzyme of the bacteriophage T7 type (Cermakian *et al.*, 1996). Unlike the *E. coli rpoC* gene, the malarial one is split into two genes ($rpoC_1$ and $rpoC_2$), as in other plastids. However, a frameshft in a poorly conserved central region of the $rpoC_1$ does not occur in other plastid genomes (Wilson *et al.*,

Figure 1.4

Ribosomal protein gene operons in *E. coli* compared with those of various plastid genomes (Wilson *et al.*, 1996).

Species include two red algae (*Porphyra purpurea*, *Cyanophora paradoxa*), the apicoplast plastid genome of *Plasmodium falciparum*, a photosynthetic protist (*Euglena gracilis*), a lower plant (*Marchantia polymorpha*), a higher plant (*Nicotiana tabacum*) and a parasitic flowering plant (*Epifagus virginiana*). Pseudogenes are indicated by open ovates and genes possibly transposed to the nucleus are indicated by unnumbered stippled ovates. Associated genes include *secY*, *rpoA* and *tufA*.


E.coli
S10
S10 L3 L4 L23 L2 S19 L22 S3 L16 L29 S17
Porphyra
L3 L4 L23 L2 S19 L22 S3 L16 L29 S17
Cyanophora
L3 L2 S19 L22 S3 L16 S17
Plasmodium
L4 L23 L2 S19 S3 L16 S17
Euglena
L23 L2 S19 L22 S3 L16
Marchantia
L23 L2 S19 L22 S3 L16
Nicotiana
L23 L2 S19 L22 S3 L16
Epifagus
L2 S19 S3 L16
E.coli
spc
L14 L24 L5 S14 S8 L6 L18 S5 L30 L15 secY L36
Porphyra
L14 L24 L5 S8 L6 L18 S5 secY L36
Cyanophora
L14 L5 S8 L6 L18 S5 secY L36
Plasmodium
L14 S8 L6 S5 ORF 91 L36
Euglena
L14 L5 S8 L36
Marchantia
L14 S8 L36
Nicotiana
L14 S8 L36
Epifagus
S8 L36
E.coli
alpha
S13 S11 S4 rpoA L17
Porphyra
S13 S11 rpoA
Cyanophora
S13 S11 rpoA
Plasmodium
S11
Euglena
Marchantia
S11 rpoA
Nicotiana
S11 rpoA
Epifagus
E.coli
str
S12 S7 fus tufA
Porphyra
S12 S7 tufA S10
Cyanophora
S12 S7 tufA S10
Plasmodium
S12 S7 tufA
Euglena
S12 S7 tufA
Marchantia
S12 S7
Nicotiana
S12 S7
Epifagus
S12 S7

1996). The intron characteristic of other plastid versions of *rpoC1*, is also absent in *rpoC1* of the malarial 35kb DNA. The plastid genome of the higher plant *Epifagus virginiana* on the other hand, contains a truncated pseudogene for *rpoA*. This is the only remnant of the *rpo* genes left on this higher plant vestigial plastid genome.

A total of 25 tRNA genes, a minimal set covering all 20 amino acids, was also identified on the 35kb plastid genome of *P. falciparum* (Preiser *et al.*, 1995). This finding contrasts with mitochondria which take up some of their tRNAs from the cytoplasm (Sugiura, 1992). Furthermore, one tRNA gene has been found which contains an intron of 130nt (tRNA $^{Leu(UAA)}$). Although smaller than the introns found in the tRNA genes of other plastids it is located in the anticodon loop of the cloverleaf structure, a conserved position for plastid tRNA introns (Preiser *et al.*, 1995). More recently, analysis of the 35kb tRNA-Ile (an isoleucine tRNA with a CAU anticodon that is post-transcriptionally modified to read AUA, (Osawa *et al.*, 1992), shows it has an unusual stem-loop in the variable region (Blanchard and Hicks, 1999). This extension has only ever been seen in the plastid tRNA of the red algae, *Cyanophora* and *Porphyra*, the diatom *Odontella*, the photosynthetic protist *Euglena* and the green alga *Chlorella*. It is absent from all other plastid, mitochondrial or bacterial tRNAs with the CAU anticodon. The presence of this insertion is indicative of a plastid ancestry. But as it is observed only in a minority of plastid genomes, the insertion must have occurred after the primary origin of the plastid, with a secondary loss taking place during the evolution of land plants consistent with the monophyly of plastid lineage.

The final piece of evidence indicating a plastid rather than mitochondrial origin for the malarial circular genome is that it does not carry typical mitochondrial genes, such as cytochrome *c* oxidase subunit I and cytochrome *b*. These cytochrome genes are observed only on the 6kb element of *P. falciparum* (Vaidya *et al.*, 1989; Feagin, 1992).

1.2.2.2 Red or green?

Further evidence of an algal plastid origin was obtained by analysis of the predicted peptide product of the open reading frame ORF470 carried on the malarial 35kb DNA. It was found to be ~50% identical with an ORF present on the plastid (rhodoplast) genomes of three red algae, *Antithamnion* sp. (Kostrzewa and Zetsche, 1992), *Cyanidinium caldarium* and *Porphyra purpurea* (Wilson *et al.*, 1993; Williamson *et al.*, 1994). This discovery suggested a rhodoplast origin for the malarial

DNA. However, phylogenetic studies comparing the plastid *tufA* gene of three apicompexan parasites (*P. falciparum*, *T. gondii* and *Eimeria tenella*) grouped the organellar genome with cyanobacteria and green algal plastids (Kohler *et al.*, 1997). In this study, consistent clustering of malarial *tufA* with green algal plastids by all analytical methods used suggested a green rather than red algal origin for the apicoplast genome. Although the *tufA* gene is a common feature of algal plastid genomes it is absent from the chloroplast genomes of higher plants. Phylogenetic analysis suggests that *tufA* was transferred from the chloroplast to the nucleus within the green algal lineage, in the evolution of land plants (Baldauf *et al.*, 1990). The position of the *tufA* gene, next to the truncated *str* operon (see Figure 1.4) is also indicative of an algal origin, like the plDNA of euglenoids.

Whether or not the 35kb DNA is of green or red algal in origin, these observations are consistent with the evolutionary evidence that apicomplexans are related to dinoflagellates (Levine, 1988; Barta *et al.*, 1991; Gajadhar *et al.*, 1991, Silberman *et al.*, 1996), many of which also contain plastids as (shown by their brown pigmentation). This evidence rests on phylogenetic analysis of small subunit rRNA genes, (16S-like) rRNAs which are widely used for estimating genetic relatedness over broad evolutionary ranges. The evolutionary relationship between dinoflagellates and apicomplexans is discussed further in section 1.4.1.

1.2.3 Questioning the evolution of Apicomplexans

The unexpected discovery of a plastid genome in apicomplexans, sparked an interest in the plant-like features of other malarial nuclear genes (Wilson *et al.*, 1994), with a view to finding the distant evolutionary relationship of apicomplexans with a photosynthetic organism. In 1993, Robson *et al.* placed the malaria parasite calcium binding protein, calmodulin, in a monophyletic group with ciliates and plants. A year later, in 1994, an indel (insertion sequence) of five amino acids was found in the enolase gene (encoding an enzyme involved in glycolysis) of *P. falciparum* (Read *et al.*, 1994). Analysis of this indel showed its position and primary sequence was apparently characteristic of that found uniquely in higher plant enolase proteins. Similarly the observation that the branch point for the malarial histone H2A gene in dendrograms lay close to that of the pine tree (Thatcher and Gorovsky, 1994), suggested an evolutionary connection with higher plants. Further evidence of a plant-like relationship was gained with the bradyzoite-specific heat-shock protein, Hsp30,

from *T. gondii*. The predicted amino acid sequence of the 687bp open reading frame encoding this gene showed similarity to the conserved C-terminus of the 18kDa heat shock proteins of plants (Bohne *et al.*, 1995). However as described by Wilson and Williamson (1997) these observations were not congruent with classical taxonomy (Levine, 1987) and molecular phylogeny of the nuclear SSU rRNA genes which relate apicomplexans to the dinoflagallate/ciliate clade. Barta *et al.* (1991) showed this for species of *Eimeria*, Escalante and Ayala (1995) with *Plasmodium* sp. and Gajadhar *et al.* (1991) showed an evolutionary relationship between the apicomplexan parasites *Sarcocystis muris* and *Theileria annulata* with the free-living dinoflagellete, *Crypthecodinium cohnii*. This fitted with the earlier observation of Taylor in 1974, who noted that the triple-layered plasma membrane of apicomplexans resembles that of ciliates and dinoflagellates. Furthermore, although the discovery of the plastid genome can be construed to indicate some plant-like ancestry, the hypothesized endosymbiotic origin of the plastid (see 1.4.1) genome fits better with the accepted dinoflagellate-related ancestry of the apicomplexans (Van der Peer *et al.*, 1996; Wilson and Williamson, 1997).

1.3 The plastid organelle

1.3.1 Discovery

Association of the circular 35kb circular plastid genome with an organelle was first indicated following genetic crosses and DNA analysis experiments on male and female gametes in malaria. In one experiment Vaidya *et al.* (1993) crossed two *P. falciparum* clones, HB3 and Dd2. Following extraction of DNA from their progeny Southern analysis was carried out using parent-specific probes to both SSU rRNA and *rpoB* genes. The results indicated inheritance was maternally dominated. In addition to this genetic cross, both Vaidya *et al.* (1993) and Creasey *et al.* (1994) found that probes for part of the *rpoB* gene hybridized with DNA from purified female (macro), but not with purified male (micro) gametes of *P. gallinaceum*. In malaria it is only the female gametes which contain a full complement of cytoplasmic organelles. Male gametes on the other hand, are comprised of a nucleus with a minimum of cytoplasm. Consequently, cytoplasmic genomes are almost exclusively inherited from the maternal parent.

In situ hybridization of tissue stages of the apicomplexan parasite *Toxoplasma gondii* has confirmed a distinct organellar location of the 35kb genome. *T. gondii*

carries up to eight copies of the 35kb circle per haploid genome (as opposed to one in *P. falciparum*). Hybridization with ring and tropohizoite stage parasites, using a digoxygenin-labelled DNA probe for a variable region of the small subunit rRNA encoded by the 35kb circle, revealed a discrete fluorescence signal. This was easily distinguished from the nucleus, and often corresponded to weak DAPI-staining of a source of extrachromosomal DNA (McFadden *et al.*, 1996). A similar experiment a year later confirmed this: fluorescence *in situ* hybridization to both plastid DNA and RNA transcripts revealed that the 35kb genome of *T. gondii* was situated adjacent to (but distinct from) the apical end of the parasite nucleus (Kohler *et al.*, 1997).

To further examine the subcellular location of the plDNA, *T. gondii* was again chosen because the regular organization of intracellular organelles and well defined apical region offered better ultrastructural resolution than *P. falciparum* (Kohler *et al.*, 1997). Frozen ultra thin sections of tachyzoites were hybridized with digoxygenin-labelled DNA probes. These were probed with antidigoxygenin antibody followed by a second antibody and gold-conjugated protein. This showed that the 35kb element was associated with a discrete organelle, adjacent to the nucleus but distinct from either the mitochondrion or the Golgi apparatus. The presence of four bi-layer membranes surrounding this organelle was determined by electron microscopy using Epon-embedded parasites (for membrane preservation). An electron micrograph of *T. gondii* highlighting the plastid organelle (double arrow) is given in Figure 1.5 (Wilson and Williamson, 1997). The presence of multiple membranes supported the hypothesis that an early progenitor of the phylum Apicomplexa acquired an algal plastid by secondary endosymbiosis (Wilson *et al.*, 1994). Plastids with two membranes (such as those from algae, plants and glaucophytes) are thought to derive from a unique, primary endosymbiosis of a cyanobacterium (Cavalier-Smith, 1995). On the other hand, secondary endosymbiosis, where a eukaryote engulfs another eukaryote containing a plastid, permanently retaining part of it as a degenerate symbiont, results in organelles with more membranes as shown in Figure 1.6 (Palmer and Delwiche, 1996).

Indeed, the multi-membraned plastid organelle had been observed in electron micrographs of *Plasmodium* species for more than 20 years (McFadden and Waller, 1997), and noted as a "spherical body" (Aikawa, 1966), a "vacuole" (Rudzinska and Vickerman, 1968) or a "double-walled vesicle" (Hackstein *et al.*, 1995). More

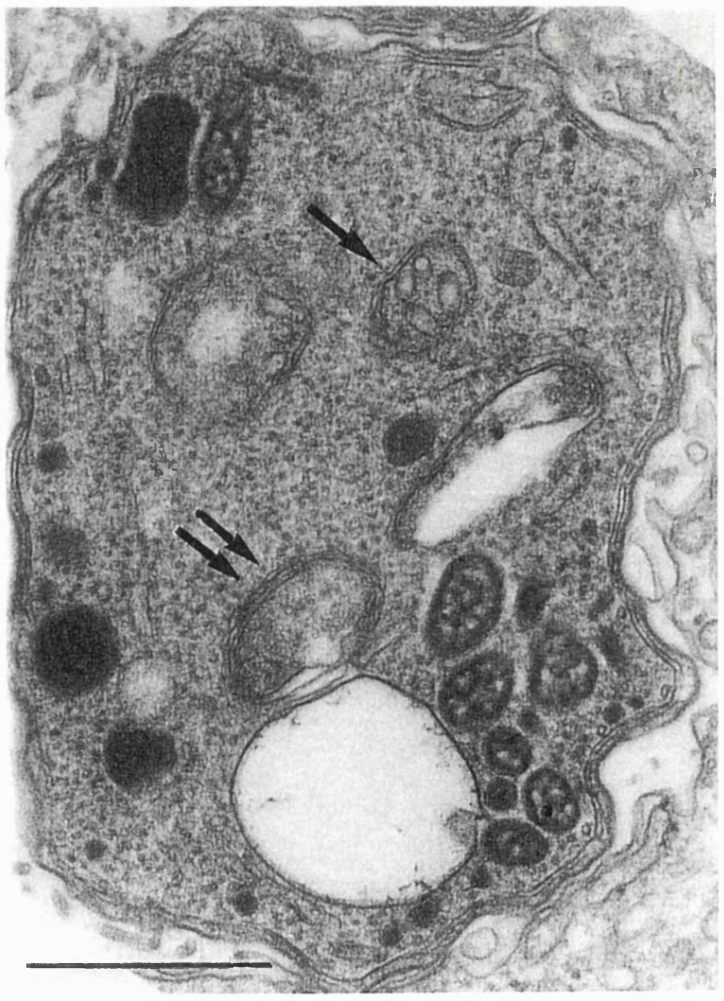


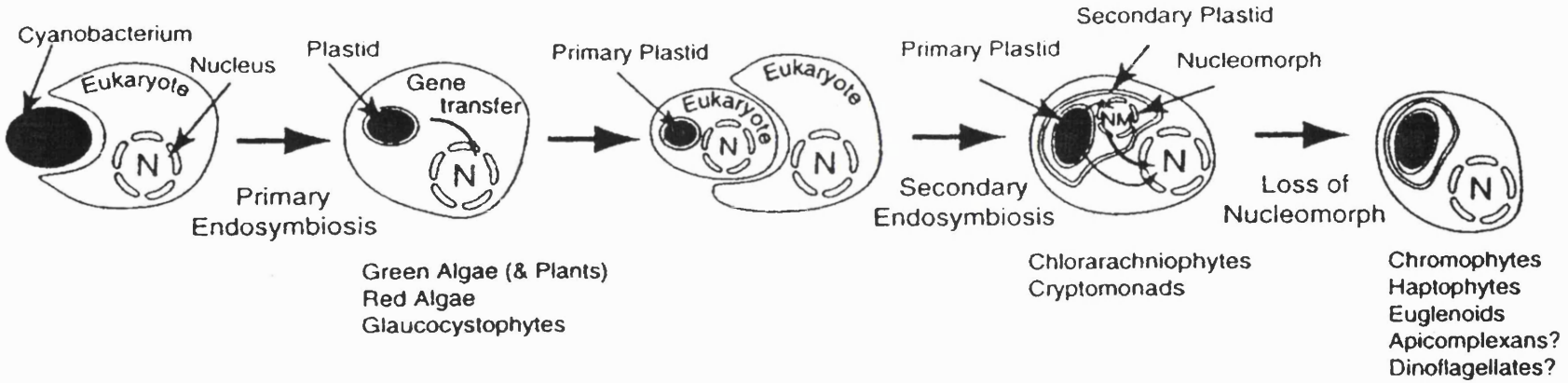
Figure 1.5

Electron micrographic section through the tubular mitochondrion (arrow) and plastid (double arrow) of a *Toxoplasma gondii* tachyzoite (Wilson and Williamson, 1997).

Figure 1.6

Origin and spread of chloroplasts by primary and secondary endosymbiosis (Palmer and Delwiche, 1996).

Groups of eukaryotes with plastids arising from either primary or secondary endosymbiosis are listed below the appropriate cell type. Secondary endosymbiosis, where a eukaryote engulfs another eukaryote (containing a plastid derived from primary endosymbiosis) results in additional membranes. Sometimes four (as shown), sometimes three (Palmer and Delwiche, 1996).


Cyanobacterium
Nucleus
Eukaryote
N
Primary
Endosymbiosis
Plastid
Gene
transfer
N
Green Algae (& Plants)
Red Algae
Glaucocystophytes
Primary Plastid
Eukaryote
N
Eukaryote
N
Secondary
Endosymbiosis
Secondary Plastid
Primary Plastid
Nucleomorph
NM
N
Chlorarachniophytes
Cryptomonads
Loss of
Nucleomorph
N
Chromophytes
Haptophytes
Euglenoids
Apicomplexans?
Dinoflagellates?

recently a three-dimensional ultastructural analysis of the plastid organelle from *P. falciparum* indicates that unlike *T. gondii*, the malarial plastid has a triple-membraned wall (also observed in *P. knowlesi* and *P. berghei*) (Hopkins *et al.*, 1999). In addition, two localized sets of more elaborate membrane whorls which increased in complexity during the intraerythrocytic life cycle were observed. The first of these two membrane complexes (seen as wrinkles in one or more of the basic three membranes) is known as the outer membrane complex and is observed between the outer and inner membranes of the malarial plastid. However, its complex configuration has not yet allowed its assignment to either the outer or middle membrane. The second (inner membrane complex) is observed within the lumen of the plastid and is thought to be a rolled invagination of the plastid's inner membrane. It is not clear why the plastid of *Plasmodium* is contained by fewer membranes than *Toxoplasma*. It may be that *Plasmodium* species lost one of the membranes during its evolutionary history or that the additional membrane in *Toxoplasma* is similar to the extra membranes seen in *P. falciparum* (Hopkins *et al.*, 1999).

As well as membrane number and information on the organelle's internal structure, the ultrastructural analysis carried out by Hopkins *et al.*, (1999) showed association of the plastid with other cellular components. In particular, it was closely associated with the mitochondrion along its whole length in merozoites and early rings, but only at one end in later stages. This would be consistent with a possible metabolic interaction between the two organelles, first suggested by Aikawa *et al.* (1967) for *Plasmodium elongatum*. Regions of the plastid were also shown to be closely associated with the pigment vacuole, nuclear membrane and endoplasmic reticulum. These associations were suggested possibly to indicate transport to and from the organelle (see Sections 1.4.3.2 and 1.5.2.2). Furthermore, in merozoites the plastid was anchored to a band of 2-3 subpellicular microtubules. This could provide organellar movement following division (Section 1.3.2). The internal matrix of the organelle was found to contain ribosome-like granules and a network of fine branched filaments. Similar filamentous structures are apparent in some mitochondria, and it has recently been suggested that these may aid the transport of DNA (accumulated by mitochondria from the cytoplasm) to the nucleus (Nass and Nass, 2000). This is due to the frequently observed contacts of mitochondria with the nuclear membranes in the early stages of chick embryo development. However the role of these filaments in

the apicoplast is unknown. A diagrammatic representation of the structure of the malarial plastid from Hopkins *et al.* (1999) is given in Figure 1.7.

1.3.2 Replication and division

Plasmodium falciparum 6kb mitochondrial DNA replication coincides with nuclear replication in the erythrocytic cycle, at or just before the onset of schizogony (Preiser *et al.*, 1996; Smeijsters *et al.*, 1994). A rolling circle activity, has been identified which is observed also in certain bacteriophages and plasmids (Preiser *et al.*, 1996). The time and replication of the 35kb genome could not be distinguished from the replication of the 6kb mitochondrial DNA. This suggests that in spite of their different location, both types of DNA are replicated in close concert (Smeijsters *et al.*, 1994). Replication of the 35kb molecule is initiated in the inverted repeat and occurs by D-loop formation, followed by rolling circle activity. Replication bubbles have been observed in both *P. falciparum* and *P. knowlesi* (P. Preiser, NIMR, personal communication).

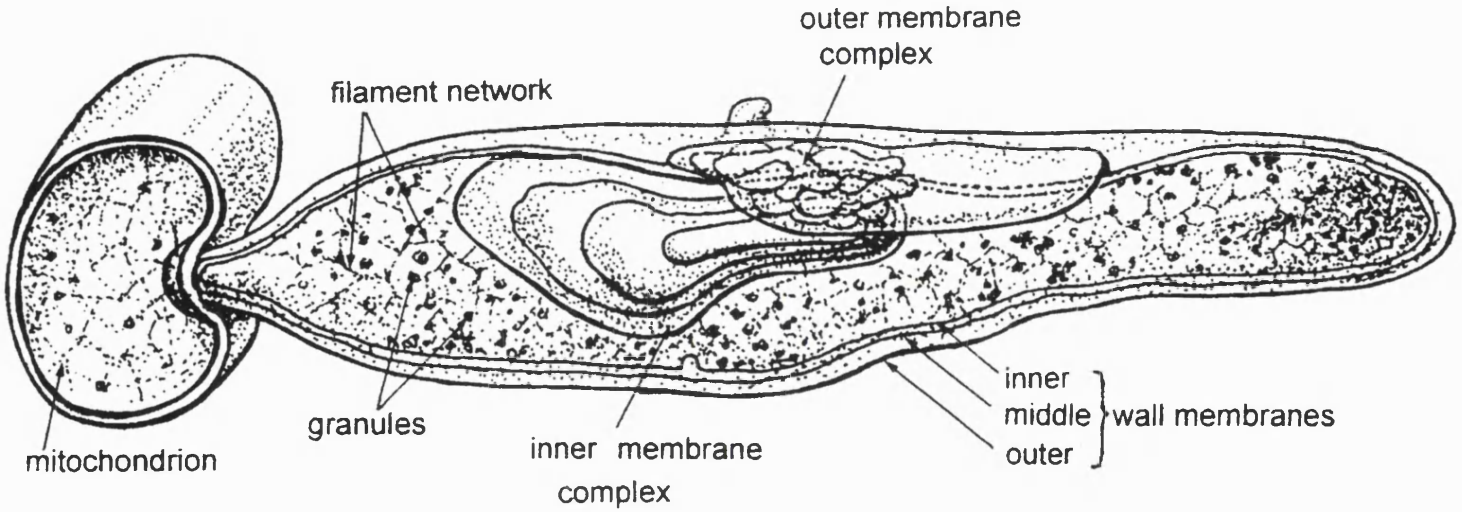
Due to the endosymbiotic origins of both mitochondria and plastids, it seemed likely that organelle kinesis would be regulated by a contractile ring. The first plastid dividing (PD) apparatus was observed in the primitive red alga, *Cyanidium caldarium*, in 1986 (Mita *et al.*). This ring is located in the cytoplasm outside the organelle envelope at the constriction point of dividing organelles. Subsequently a PD ring has been observed in all eukaryotic plants examined. Eukaryotic cell division therefore is controlled by at least three division apparati: a contractile ring, a MD (mitochondrial-dividing) ring and a PD (plastid-dividing) ring. Conversely, bacterial division is controlled by a single division apparatus: the FtsZ ring (Kuroiwa *et al.*, 1998).

Division of the apicoplast has been described for *T. gondii* (Streipen, http://www.sas.edu~streipen/gfp.html; McFadden *et al.*, 1996) and *P. falciparum* (Streipen, http://www.sas.edu~streipen/gfp.html; Hopkins *et al.*, 1999) however, the details are rather vague. Electron micrographs of *T. gondii* (McFadden *et al.*, 1996) showed only one organelle per cell, except during division (endogeny) when an organelle was associated with each daughter cell. At the onset of endogeny the plastid organelle was seen to assume a cylindrical shape, eventually acquiring a constriction and a characteristic dumbbell shape. This was then assumed to divide into two. A single plastid was also observed for *P. falciparum* until its replication in late

Figure 1.7

Reconstruction of the malarial plastid organelle from serial sections (Hopkins *et al.*, 1999).

The organelle has two localized and elaborate outer and inner membrane complexes. The interior shows the granular structure and system of branched filaments. The indentation of the wall of the mitochondrion indicates the close association between plastid and mitochondrion.


outer membrane
complex
filament network
inner
middle
outer
wall membranes
granules
inner membrane
complex
mitochondrion

schizogony when the numbers increased up to 16 (Hopkins *et al.*, 1999). Recent studies (using a green flourescent protein reporter construct) have shown that in *P. falciparum*, single apicoplasts enlarge slowly before elongating into branched forms during schizogony. These then divide and segregate prior to merozoite release; each new merozoite inheriting only one plastid organelle (Waller *et al.*, 2000). The plastid in *P. falciparum* has also been associated with microtubules which may be involved in transport of newly divided plastids (and its closely associated mitochondrion) into assembling merozoites, although this has not yet been proven (Hopkins *et al.*, 1999).

1.4 Endosymbiosis and beyond: the acquisition and maintenance of a vestigial plastid genome in malaria

1.4.1 The "Mr Gobbler" hypothesis

Although the malarial plastid genome, contained within a distinct multi-membraned organelle was an unexpected discovery, the endosymbiotic algal origin fits well with the accepted dinoflagellate-related ancestry of the apicomplexans described above (Wilson *et al.*, 1994).

Dinoflagellates are a diverse, abundant group of aquatic unicellular protozoa of ancient origin. Fossil traces of their siliceous envelopes have been found as far back as the Precambrian/Cambrian transition (Escalante and Ayala, 1995). Some species are photosynthetic and have a brownish coloration due to their plastids (Cavalier-Smith, 1993). The plastids of the dinoflagellates are thought to have been acquired by their habit (or that of a dinoflagellate progenitor) of capturing algae. Furthermore, by analysing chloroplast genes, Zhang *et al.* (1999) showed that dinoflagellates, whose plastids have three membranes, probably engulfed a red-algal-like plastid. Following acquisition of an alga, substantial transfer of algal genes to the nucleus of the host cell has to occur to allow maintenance of plastids as organelles (Gibbs, 1981). Wilson and Williamson (1997) suggested this secondary endosymbiosis event served the dinoflagellate well until 9 x 10^8 years ago when a single progenitor cell became parasitic within a polychete worm (Levine, 1987). The plastid then (if not before) lost its photosynthetic capability and was retained only for some secondary activity. The suggestion that polychete worms may have been an important intermediate host in the evolution of the apicomplexans is interesting as they were ancestors of the Insecta (and mosquito). Furthermore it is thought that apicomplexans like *Plasmodium* were probably well adapted to their insect vectors

long before vertebrate hosts appeared (Escalante and Ayala, 1995). Very little is known about the early apicomplexans, but a relic plastid occurs in representatives of all three major apicomplexan lineages (McFadden *et al.*, 1997); the haemosporins (*Plasmodium*), the piroplasms (*Babesia* and *Theileria*) and the coccidians (*Toxoplasma*, *Eimeria* and *Sarcocystis*). Portions of the plDNA (in particular the inverted ribosomal repeat) from four apicomplexan genera covering about one third of the total length of the *Plasmodium* plDNA were shown to have a highly conserved gene content, order and sequence. As the plastid genome is so reduced it would have required many deletions and rearrangements to occur throughout its evolution. Thus, the high level of similarity amongst the apicomplexan genera suggests a single photosynthetic progenitor (Denny *et al.*, 1998). The "Mr Gobbler" hypothesis of how the malaria parasite acquired its plastid is shown in Figure 1.8 (Wilson *et al.*, 1994).

1.4.2 Translocation of plastid-encoded genes

1.4.2.1 The fate of "promiscuous" DNA

There is extensive evidence for the movement of DNA across cellular membranes inside eukaryotic cells. The compelling evidence for an endosymbiotic origin of both mitochondria and chloroplasts (Cedergren *et al.*, 1988; Gray, 1989), coupled with the small size of their present genomes (Palmer, 1990) implies that genetic material must have been lost from these organelles. Plastid genomes generally contain only ~1-5% the number of protein-coding genes on present day cyanobacterial genomes. For example, the unicellular cyanobacterium *Synechocystis* encodes 3168 proteins (Martin and Herrmann, 1998), whereas most plant chloroplasts contain between 60 and 80 protein-coding genes, and the rhodophyte *Porphyra* sp. has ~200. Likewise *Odontella* sp. and *Cyanophora* sp. have between 120 and 130 genes. Some of this "promiscuous DNA" (Ellis, 1982; Cerutti and Jagendorf, 1995) has ended up translocated to the nucleus (Baldauf *et al.*, 1990; Gantt *et al.*, 1991; Nugent and Palmer, 1991). However, it can also relocate in the mitochondrial genome (Ellis, 1982). This phenomenon, as well as the transfer of nuclear sequences to the mitochondrial genome has been observed in several species, eg. maize (Stern and Lonsdale, 1982) and *Oenothera* (Schuster and Brenicke, 1987). The translocation of DNA, particularly to the nucleus, has resulted over time in its concentration in nuclear chromosomes, culminating in a compartmentalized, but integrated, eukaryotic genetic

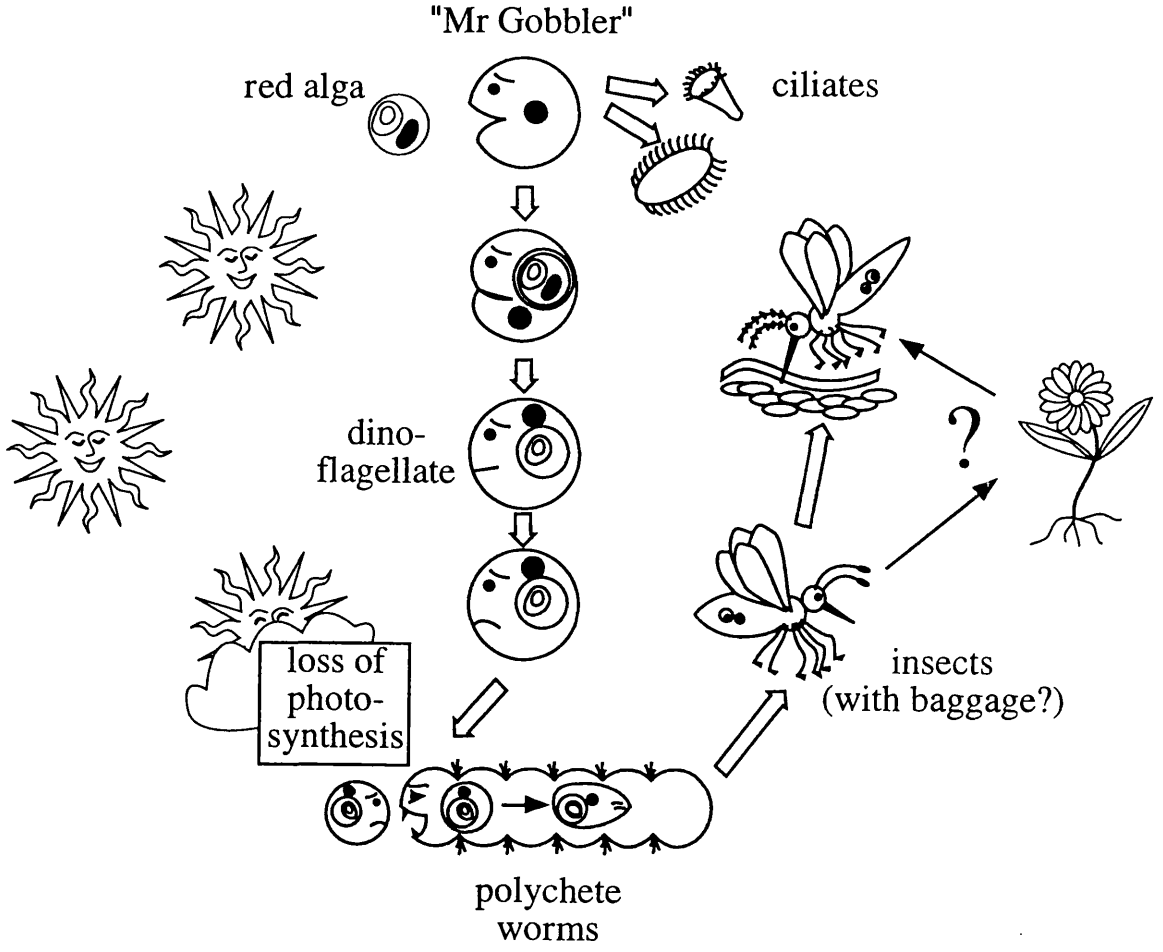


Figure 1.8

How the malaria parasite got its plastid DNA – the "Mr Gobbler" hypothesis (Wilson *et al.*, 1994).

A dinoflagellate (or dinoflagellate progenitor) acquired an algal plastid by secondary endosymbiosis. This was followed by transposition of algal genes to the nucleus. Later a dinoflagellate became parasitic within a polychete worm, and the plastid (if not before) lost its photosynthetic capabilities, being retained for some secondary activity. Evolution of polychete worms (containing an obligate intracellular parasite) into Insecta followed. These early apicomplexan parasites within their insect hosts then adapted to include a wide variety of host organisms, including humans.

system under the regulatory dominance of the nucleus (Herrmann, 1997). Although intracellular gene transfer is an on-going process, most genes were transferred during the early phases of endosymbiosis (Baldauf *et al.*, 1990; Gantt *et al.*, 1991). However, the number of nuclear genes known to descend from cyanobacterial genomes is unknown (Martin and Herrmann, 1998).

1.4.2.2 What type of genes have been lost to the nucleus and why

Cataloguing the genes carried by cyanobacteria (the ancestor of plastid genomes as shown in Figure 1.6) with those retained by plastid genomes has been used to determine what type of genes have been relocated. In higher plants, genes for phycobilisomes and eubacterial cell wall synthesis have been lost all together. However, Herrmann (1997) suggests that genes for regulatory functions tend to be fixed in the nucleus more readily than those specifying enzymatic or structural functions. In accordance with this theory, most chloroplast genes are involved in translation and photosynthesis, whereas in mitochondria they are those involved in respiration and translation.

Although no longer plastid-encoded, many of the original cyanobacterial functions still exist in plastids, such as amino acid biosynthesis (Martin and Herrmann, 1998). In fact it has been suggested (also by Martin and Herrmann) that more than 2000 proteins may be found in the average plastid, far exceeding the number of plastid-encoded genes. This indicates that plastids import the vast majority of their proteins, which are encoded by the nucleus. So during evolution the products of genes that have been relocated from organelles (mitochondria and plastids) have to be re-imported back into the appropriate organelles.

Why would organellar genes transfer to the nucleus? One suggestion is that redox-associated functions in organelles increased the chance of free-radical-induced mutations in their genomes. Protection from such mutations would favour gene transfer to the nucleus (Allen and Raven, 1996).

A second suggestion is that when the plastid originally entered its host cell by endosymbiosis it became asexual, having been isolated from its free-living relatives. This resulted in the loss of genetic recombination through sexual reproduction which would allow any deleterious mutations to be recombined out of its genome. Muller's ratchet hypothesis (Muller, 1964) stated that given sufficient time asexual populations and species would become extinct because of accumulating mutations. Therefore it is

hypothesized that gene transfer to the nucleus allowed recombination to occur again, i.e. resulted in the resumption of a predominantly sexual lifestyle from a predominantly asexual one.

Evidence for Muller's ratchet has been seen in organelle genomes, in particular the tRNA genes of animal mitochondria (Lynch, 1996). However, the effect of Muller's ratchet is much less pronounced in the organelles of higher plants. In fact Wolfe *et al.* (1987) found the rate of nucleotide substitution lower in mitochondria and chloroplasts than the nucleus. This suggested that compensatory factors were counteracting the long-term affects of asexuality. In *Chlamydomonas* this occurs (Fischer *et al.*, 1996) by recombination between chloroplast organelles. Also, recombination between genomes within the same plastid was indicated as plastids are known to import a nuclear encoded homologue of cyanobacterial RecA, a protein involved in recombination and repair (Cerutti *et al.*, 1995). Other DNA repair pathways which may help to account for the lower substitution rate include excision repair. Although this has not yet been characterized in plastids, one gene from this pathway (*mutS*) has been found in the animal mitochondrial genome (Pont-Kingdon *et al.*, 1995).

Although Muller's ratchet theory cannot account for all the transfer of DNA from organelles into the nucleus, it may have played an important part in early evolutionary history, i.e. before recombination between organellar DNA and organellar repair pathways evolved (Martin and Herrmann, 1998).

The apicomplexan plastid differs from that of true plants and algae because it is derived from a secondary endosymbiotic event, resulting in a multi-membraned organelle. Figure 1.6 shows that after the primary endosymbiotic event, gene transfer must first have taken place from the cyanobacterium to the algal nucleus. Following secondary endosymbiosis the nucleomorph then "fused" with the nucleus, in addition to further gene transfers from the plastid genome.

1.4.2.3 Evidence for translocation to the nucleus in secondary endosymbionts

Gene transfer from an unrelated organism containing a plastid of secondary endosymbiotic origin has been studied in the photosynthetic protist, *Euglena gracilis*. Two examples of nuclear gene-encoded chloroplast-targeted enzymes of *Euglena* are those for ribulose-bisphosphate carboxylase small subunit (*rbcS*) and light harvesting chloroplast protein I (LHCI), (Chan *et al.*, 1990; Houlne and Schantz, 1988). In addition Henze *et al.*, (1995) cloned the cDNA of a cytosolic GAPDH (glyceraldehyde-3-

phosphate dehydrogenase) enzyme, GapC, an enzyme that is involved in cytosolic glycolysis. The phylogeny of these sequences indicated gene transfer from eubacteria via organelle-to-nucleus transfer.

Through Southern blot analysis, Hackstein *et al.* (1995) showed that *psbA* sequences were present in the genomic DNA of two apicomplexan parasites, *T. gondii* and *Sarcocystis muris*. However, hybridization to DNA from *P. falciparum* was weak. *pbsA* codes for the DI protein of the photosystem II reaction centre of thylakoid membranes and is involved in photosynthesis. The *psbA* gene was subsequently isolated from the merozoites of *S. muris* by PCR and sequence analysis confirmed the product to be *psbA*. The putative polypeptide, from the deduced amino acid sequence, had conservation of more than 90% in the triazine herbicide-binding region. Triazine is effective in blocking photosystem II and Baycox (a modified triazine derivative) has been shown to be affective on apicomplexans. Mehlhorn *et al.* (1984) and Harder and Habercorn (1989) showed this in *Eimeria falciformis* and *E. tenella* as did Hackstein *et al.* (1995) in *T. gondii*. High performance liquid chromatography analysis on extracts of *T. gondii* trophozoites (Hackstein *et al.*, 1995) showed they also contained photochlorophyllidea *a* (an intermediate of chlorophyll biosynthesis) and chlorophyll *a*. The latter was indicated by subjecting frozen cells to 77K fluorescence spectroscopy which yielded an emission spectrum with maximum fluorescence at 695-689 and 716nm. In photosynthetic organisms this indicates chlorophyll *a* bound to PSI and PSII (Rijgersberg *et al.*, 1979; Rijgesberg and Amesz, 1980), leading Hackstein *et al.* to hypothesise that apicomplexans harbour a chlorophyll *a*-D1 complex within the plastid organelle which must be coded for by nuclear genes. However, this has never been substantiated. Furthermore, the observation that triazines also inhibit some non-plastid containing organisms (Harder and Habercorn, 1989), suggests that these compounds have an alternative target other than the DI protein. Nevertheless, the discovery of nucleus-encoded protein import into the apicoplast (Waller *et al.*, 1998) suggests that translocation has occurred, and is discussed below.

1.4.3 Importation of nuclear-encoded proteins into plastids

As mentioned previously, many of the original cyanobacterial functions still exist in plastids which require proteins translated from plastid-encoded genes translocated to the nucleus (Martin and Herrmann, 1998). Furthermore, the absence of numerous essential housekeeping genes from the apicomplexan plastid genome such as those for

FtsZ, EF-G and ClpP indicates that protein transport must be occurring in apicomplexans, just to allow maintenance of the plastid. The first of these, FtsZ is an essential component of prokaryotic cell division, being responsible for the formation of the contractile ring. FtsZ is also thought to be involved in the division of chloroplasts (Lutkenhaus, 1998). EF-G, on the other hand, is one of the missing protein synthesis components of the plastid genome. It is required alongside EF-Tu to recycle GTP during the elongation cycle of protein synthesis in prokaryotes and chloroplasts (Nierhaus, 1996). Furthermore, in soybean (*Glycine max*), EF-G is a nuclear encoded protein targeted to the chloroplast (Hernandez-Torres *et al.*, 1993). The other housekeeping gene mentioned above, *clpP*, encodes an ATP-dependent serine protease (Maurizi *et al.*, 1990*a*,*b*), which associates with a nucleus encoded regulatory subunit, ClpC as a complex. ClpP is chloroplast encoded in plants and in algae both proteins are chloroplast encoded. On the apicoplast genome a gene encoding ClpC was identified (Wilson *et al.* 1996). As both subunits have been localized in the stroma of plant chloroplasts (Shanklin *et al.*, 1995), it is likely that ClpP (encoded on chromosome 3 of *P. falciparum*, Bowman *et al.*, 1999) is imported into the plastid of apicomplexans in order to associate with ClpC.

1.4.3.1 Chloroplast import

The import of precursor polypeptides into chloroplasts is becoming increasingly well understood (Cline and Henry, 1996; Keegstra and Cline, 1999). The process can be broken down into three distinct phases: specific targeting of precursor proteins to plastids, transport across the two envelope membranes, and in many cases sorting to the proper plastidic compartment (Keegstra and Cline, 1999).

Following the expression of nuclear encoded genes, immature precursor polypeptides are synthesized in the cytosol with an N-terminal transit peptide, which is responsible for routing to the correct organelle (Doberstein *et al.*, 1977). Despite their common function, no consensus targeting sequences have been established, nor have common secondary structures been observed. However, the N-terminal chloroplast-directing targeting sequences have some features in common. These include a high serine and threonine content, an uncharged N-terminal region with an overall positively-charged transit sequence and primary and secondary motifs next to the cleavage site (von Heijne *et al.*, 1989; Bartling *et al.*, 1990; von Heijne and Nishikawa, 1991). This N-terminal sequence usually contains stromal-targeting information, allowing the transport

of precursors across the two envelope membranes (Keegstra, 1989). In some, cases the transit peptide is bipartite and contains additional targeting information. Precursor proteins requiring this additional information are those that require transportation to the thylakoid lumen. Proteins directed towards the outer envelope of the chloroplast lack N-terminal targeting sequences (Soll and Alefsen, 1993), information for targeting being contained within the mature protein (Cline and Henry, 1996). Similarly, precursors destined for insertion into internal chloroplast membranes generally contain target information within the mature region of the peptide rather than within additional transit peptide sequences.

Import into the chloroplast is initiated by binding of the chloroplast-targeting domain to receptors on the surface of the organelle that are susceptible to protease treatment (Cline *et al.*, 1985). However, this treatment only restricted the binding of specific polypeptides to specific binding sites. For example pSSU (precursor of the small subunit of ribulose-1, 5-bisphosphate carboxylase oxygenase, Rubisco) protein import was inhibited by treatment with phospholipase C (Kerber and Soll, 1992). This suggested that import of components which bind with low affinity could be mediated by other import machinery (Friedman and Keegstra, 1989). One example of chloroplast proteins lacking specific receptors is those directed towards the outer envelope of the chloroplast (Soll and Alefsen, 1993).

Chaperones are thought to be the driving force for protein import (Keegstra and Cline, 1999), allowing translocation of protein precursors through the outer and inner envelope membranes of chloroplasts (Cline *et al.*, 1985). ATP hydrolysis is required for translocation of the immature protein into the inner envelope or stroma. Higher concentrations of ATP are required for transfer across the thylakoid membranes, involving both ATP hydrolysis as well as a proton motive force (Cline *et al.*, 1989). Once inside the organelle, the precursor proteins are processed by a stromal processing peptidase to remove the N-terminal signal sequences (Smith and Ellis, 1979). Precursor proteins containing multiple transit peptides require further processing into their mature forms. An example is the plastocyanin precursor, which requires secondary processing by a thylakoid peptidase (Kirwin *et al.*, 1987). Finally the mature proteins are folded in their proper organellar subcompartment (Smeekens *et al.*, 1990).

Although most research has been done on chloroplasts, many different types of plastid are competent for precursor protein uptake. These include etioplasts (Schindler and Soll, 1986), leucoplasts (Halpin *et al.*, 1989), and proplastids.

1.4.3.2 Protein import into the apicomplexan plastid

The details of protein trafficking to plastids containing multiple membranes is not well understood. However, as the outermost membrane is part of the host's endomembrane system, targeting is assumed to commence via the secretory pathway into the endoplasmic reticulum, courtesy of a classic signal peptide (Bodyl, 1997). Subsequent targeting across the inner pair of membranes involves a downstream transit peptide.

Nuclear-encoded gene-product targeting to a tri-membraned plastid organelle resulting from secondary endosymbiosis has been observed for LHCI (light harvesting-chloroplast protein I) and *rbcS* (encoding the small subunit of ribulose-bisphosphate carboxylase) of *Euglena*. Both are translated as precursor polypeptides with N-terminal and transit sequences whose subunits are proteolytically processed to yield mature proteins subsequent to transport across the three membranes (Chan *et al.*, 1990; Houlne and Schantz, 1988).

Waller *et al.* (1998) used sequence information to investigate targeting to the apicoplasts of both *T. gondii* and *P. falciparum*. Candidate genes were chosen according to their similarity to nuclear-encoded, plastid targeting sequences from those of plant and algal systems (Martin *et al.*, 1998). This allowed the identification of two ribosomal protein genes, *rps9* and *rpl28* (encoding S9 and L28) and three fatty acid biosynthetic genes, *acpP*, *fabH* and *fabZ* (encoding ACP, β-ketoacyl-ACP synthase III and β-hydroxacyl-ACP dehydratase, respectively). Molecular phylogenies confirmed these were derived from the endosymbiont (either from the plastid genome or the nucleomorph).

The five nucleus-encoded apicoplast genes encode substantial N-terminal extensions. These included an extreme N-terminal region of between 16 and 34 amino acids, thought to resemble classic signal peptide sequences containing a hydrophobic domain, followed by a "von Heijne" cleavage site (Nielsen *et al.*, 1997). Downstream of the signal peptide sequences in the *Toxoplasma* proteins, regions were identified which suggested transit peptides, i.e. they carried a net positive charge and were enriched in serine and threonine residues. The transit peptide sequences of *P. falciparum* were shorter than those of *T. gondii* (30-40 amino acids in length as opposed to 57-107 amino acids) and tended to be rich in lysine and aspargine. This was attributed to the strong A/T bias of the *Plasmodium* genome (Waller *et al.*, 1998). A recent study has shown the

serine and threonine residues are not essential components of the *P. falciparum* transit peptides (Waller *et al.*, 2000).

Localization of two of the apicoplast protein candidates (S9 and ACP) within the organelle of *T. gondii* was shown using labelled antibodies, and little or no signal was found elsewhere in the parasite. In addition to this, GFP (green fluorescent protein) fusions of *acpP* also indicated apicoplast specific targeting (Waller *et al.*, 1998).

Processing of the precursor proteins was shown by western blotting (Waller *et al.*, 1998). Both S9 and ACP antisera recognized bands of the expected sizes for the mature protein (i.e. minus the leader and signal peptides). However, larger products were observed which corresponded to the precursor peptide including the transit peptide but lacking the signal sequence. These were assumed to be processing intermediates, either crossing the multiple membranes or awaiting proteolytic cleavage in the apicoplast stroma.

Physical evidence suggesting how precursor proteins of *Plasmodium* sp. may be conveyed to the plastid from their sites of synthesis, in the endoplasmic reticulum and cisternae of the nuclear envelope, has been provided by Hopkins *et al.* (1999). Ultrastructural analysis of erythrocytic stage parasites showed close association of the outer membrane (including its leaf-like extensions) of the apicoplast with cisternae of both the nuclear envelope and endoplasmic reticulum (ER). In some places these associations looked to be continuous with the plastid's outer wall membrane (with the ER in particular) and have been hypothesized to provide a route for precursor proteins entering the plastid. It is not until very recently, however, that Waller *et al.* (2000) demonstrated protein targeting to the plastid of *P. falciparum* using GFP as a reporter molecule. Furthermore, mutational studies revealed the functional domains required for routing to the apicoplast.

The present evidence for import into apicomplexan plastids suggests that targeting is by bipartite N-terminal leaders aided by plastid, nuclear and ER membrane associations. Cleavage of the signal peptide occurs upon entry into the endoplasmic reticulum and subsequent targeting across the inner plastid membranes is facilitated by downstream transit peptides. However, the precise molecular mechanisms are not understood.

The bipartite N-terminal targeting model of Waller *et al.* (1998) has allowed the discovery of other plastid-targeted gene candidates. These include a nuclear-encoded gene for ClpP (a plastid protease mentioned above), EF-TS (elongation factor TS) and

Cpn60 (a plastid chaperonin) (Waller *et al.*, 1998). Many more genes containing putative plastid-targeting peptide sequences are likely to be identified as the malarial genome sequencing project continues.

1.5 Function

1.5.1 Is the plastid essential?

Analysis of the 35kb plastid genome of *P. falciparum* has indicated the plastid genome has been retained to fulfil an important, if not essential, cellular role. This has included the observation of translation, selective gene retention and sensitivity to antibiotics. Each of these is discussed below.

1.5.1.1 Active transcription, translation and the tRNA genes

The 25 tRNA genes, encoded by the plDNA of *P. falciparum* and covering all 20 amino acids, are believed to be sufficient to provide a potentially complete translation system. Preiser *et al.* (1995) showed that the genes, by the effective use of the wobble position, in combination with the frequency of certain codons are able to decode all codons used. In accordance with this hypothesis Preiser *et al.* proposed that Leu (UAG) can decode the four codons (CTN). The same was said for Pro (UGG), Thr (UGU), Val (UAC), Arg (UCG) and Ala (UGC) which could decode CCN, CAN, GTN, CGN and GCN respectively. In all cases a U is found at the wobble position of the tRNA (Steinberg *et al.*, 1993). This has also been observed by Mitra *et al.* (1977) in relation to the Val codon and by tRNA mutational analysis (Claesson *et al.*, 1995).

Northern blots, primer extensions, RNase protection and RT-PCR have been used to show that the genes on the plastid are transcribed and Wilson *et al.* (1996) suggested that a minimum of four transcripts were expected, all commencing within the inverted repeat. For example, northern blots and primer extension on total RNA isolated from *P. falciparum* showed that all the tRNA genes were transcribed and some appeared to be transcribed polycistronically, in accordance with the tight spacing between individual genes (Preiser *et al.*, 1995). The rRNA genes also were found to be transcriptionally active; northern blots of total parasite RNA detected transcripts of the expected sizes for both the small and large subunit rRNAs (Gardner *et al.*, 1991*a*; 1993). Also, northern blots showed that *rpoB* exists in the form of a large transcript, significantly bigger than the coding region of the gene (Gardner *et al.*, 1991*b*; Feagin and Drew, 1995).

Feagin and Drew (1995) provided evidence for stage-specific transcription of genes from the 35kb genome of *P. falciparum* using northern blot analysis. Total RNA isolated from different stages in the erythrocytic cycle of the parasite was probed for RNA polymerase subunit and SSU rRNA transcripts. RNA polymerase transcripts were almost undetectable in ring stage parasites. However, more transcript was observed in all later stages examined (trophozoite, late trophozoite/schizont and late schizont). SSU rRNA transcripts on the other hand, were least abundant in early trophozoites and most abundant in schizonts. The different patterns of transcription were attributed to RNA polymerase subunits being required for transcription of the other genes carried by the 35kb plDNA. In addition to active transcription, translation of the plDNA into proteins was first indicated by the discovery of distinct "plastid" polysomes. Wilson *et al.* initially reported this in 1996 following hybridization studies on RNA isolated from puromycin-treated and fractionated microsomes. It was observed that 0.2μM puromycin shifted the hybridization signal from the lower third of the gradients to the top, as expected for the dissociation of polysomes into ribosomal subunits. Puromycin inhibits chain elongation because its structure closely resembles that of the 3′ end of a charged tRNA molecule, causing competitive inhibition with the normal entry of AA~tRNA precursors. Thus the puromycin-terminated nascent chains fall off the ribosomes, producing incomplete polypeptides of varying lengths (Watson *et al.*, 1988). When total ribosome preparations were pre-treated with the antibiotic anisomycin, which binds only to eukaryotic ribosomes and prevents puromycin incorporation (Gale *et al.*, 1981), the puromycin shift was still observed in the RNA fractions hybridized to DNA probes (Roy *et al.*, 1999). Similar results were also seen using probes to mRNA, providing further evidence for a subset of polysomes carrying plastid rRNAs and mRNAs and indicating expression of the malarial plastid genome.

In addition to *P. falciparum* plastid protein synthesis, transcription of the plastid genome of *T. gondii* has also been indicated. *In situ* hybridization was used to localize transcripts of a plastid-like 16S rRNA which were observed by electron microscopy. Structures were highlighted, within a multi-membraned organelle, which were comparable in size to the 70S ribosomes of plastids, mitochondria and bacteria (McFadden *et al.*, 1996; 1997).

Active transcription of the genes carried by the 35kb plastid genome of *P. falciparum* as well as the observation of plastid polysomes and ribosome-like granules (described above) provides very strong evidence that the organelle has some cellular role. It is unlikely that the plastid would be carrying out protein synthesis for no reason hence it must have some important function. However, protein synthesis may be occurring purely for self maintenance thereby providing a segregating, subcellular compartment in which a vital process occurs (McFadden and Waller, 1997).

1.5.1.2 Selective retention of genes across diverse genera of apicomplexans

Having lost all of its photosynthetic and biosynthetic genes, the plastid genome of *P. falciparum* is much more reduced than that of higher plant chloroplasts. However, genes involved in gene expression have been retained not only in *Plasmodium* but also other apicomplexans. ORF470, has been maintained in apicomplexans despite being very divergent in sequence (Wilson *et al.*, 1996). This skewed distribution of genes suggests they have been retained for a reason.

Further evidence for a functional genome can be gained by comparing it to hydrogenosomes. These are bounded by double membranes and produce ATP in amitochondriate protists; they descended from the same symbiont as mitochondria, but none have been found possessing a genome (Martin and Muller, 1998). In these organelles the entire genome has translocated to the nucleus following endosymbiosis. As mentioned in section 1.4.2.2, the last mitochondrial genes to be lost from the organelle are those for translation and respiration. Therefore, it has been proposed that as no respiration occurs in hydrogenosomes, there is nothing left to translate, and hence the genes required for translation (tRNAs and rRNAs) have become unnecessary and been lost (Herrmann, 1997). This scenario suggests that the plastid genome of apicomplexans, although very degenerate, has been retained for translation and a cellular function.

1.5.1.3 Sensitivity to antibiotics

Sequence analysis of the SSU of rRNA carried out by Gardner *et al.* (1991*a*) indicated a tetracycline binding site. Not only did this support an organellar origin but antibiotic sensitivity suggested that the SSU rRNA was functional (Kiatfuengfoo *et al.*, 1989).

Further evidence that components of the malarial plastid organelle could be preferentially targeted by drugs was provided by Clough *et al.* (1997) using thiostrepton, an inhibitor of peptide bond formation and translocation (Cundliffe, 1986). The high affinity binding site for thiostrepton, A1067 in *E. coli*, in the GTPase domain of LSU rRNA, was found to be conserved in the plastid rRNA but not in the nuclear and mitochondral genomes. Inhibition of protein synthesis by thiostrepton was shown by measuring the uptake of radiolabelled amino acids in malarial cultures containing the antibiotic. *In vitro* mutational studies reduced preferential targeting of thiostrepton by introducing mutations into the *P. falciparum* pllsrRNA at position A1067 (*E. coli*). Mutation of A to U reduced binding to only ~14% that of wild type, whereas mutation of A to G reduced binding to ~35% of WT. Antibiotics inhibiting the plastid EF-Tu (encoded by *tuf*) also have antimalarial activity in blood cultures (Clough *et al.*, 1999).

The action of plastid-specific antibiotics on another apicomplexan parasite, *T. gondii*, also indicates that the plastid organelle is an essential cellular feature. Fischera and Roos (1997) showed that replication of the parasite was blocked using ciprofloxacin, a fluoroquinolone antibiotic known to affect the A subunit of bacterial DNA gyrase (Lewis *et al.*, 1996). DNA gyrase is required for replication of the plDNA. Treatment of intracellular tachyzoites with ciprofloacin resulted in a reduction of the copy number of the plastid genome (>5 copies per haploid nuclear genome) by more than ten fold over the course of replication within the infected cell. This was further reduced when the parasites lysed out of the host cell to re-infect new ones (delayed-death-phenotype).

1.5.2 What might the plastid do?

1.5.2.1 General functions of plastids

The function of the malarial plastid organelle remains a mystery, as it does for all apicomplexans. In most plants and algae, the chloroplast has a unique cellular role: photosynthesis. However, conversion from an autotrophic lifestyle to a parasitic one has occurred several times during plant evolution. Some of these plants have totally lost their photosynthetic genes but have retained their plastids in a reduced form (Wolfe *et al.*, 1992). Although the highly derived malarial plastid genome of *P. falciparum* most resembles that of the non-photosynthetic parasitic flowering plant *Epifagus*, the deletions in the apicomplexan plastid are different to those in *Epifagus*.

The deletion of genes coding for the RNA polymerase subunits, some of the tRNA genes and ORF470 in *Epifagus*, which have been retained throughout apicomplexans, implies their plastids carry out different cellular functions (Preiser *et al.*, 1995). Two functional *Epifagus* plastid-encoded genes which are absent from the malarial plastid genome are *accD* and *clpP*. The former encodes the plastid homologue of the β subunit of the carboxy transferase component of *E. coli* acetyl-CoA carboxylase (Li and Cronan, 1992), catalysing the first step in fatty acid synthesis. The other gene, *clpP*, encodes the plastid homologue of the ATP-dependent protease (ClpP) of *E. coli* (Goldberg, 1992).

Plastids are the site for cellular biosynthesis of porphyrin (needed for haem biosynthesis), essential amino acids, essential fatty acids (Boyer *et al.*, 1989; Howe and Smith, 1991; Wallsgrove, 1991), tetrapyrroles, isoprenoids, pyrimidine nucleotides, vitamin B1, the reduction of nitrite, starch metabolism and glycolysis. At least some of these functions make plastids indispensable to plants and algae, so perhaps this is the role of the apicoplast too (Wilson *et al.*, 1996; Kohler *et al.*, 1997; McFadden *et al.*, 1997; McFadden and Waller, 1997). Although essential haem and amino acid biosynthesis in the apicoplast cannot be discounted, available evidence suggests these functions are fulfilled by other cytosolic pathways (Wilson *et al.*, 1996; Kohler *et al.*, 1997; Mcfadden *et al.*, 1997; Wilson and Williamson, 1997). In the case of haem synthesis for example, an alternative pathway is already used in the mitochondrion of the malarial parasite (Surolia and Padmanaban, 1992). The cases for and against the apicoplast as a site of fatty acid and amino acid biosynthesis are given below.

1.5.2.2 Fatty acid biosynthesis?

As discussed in section 1.4.3.2, three candidate genes involved in fatty acid biosynthesis (*acpP*, *fabH* and *fabZ*) have been identified in the nuclear genome of both *T. gondii* and *P. falciparum* and carry characteristic plastid import N-terminal signal sequences. The product of one of these genes, *acpP*, has also been shown to localize to the apicoplast (Waller *et al.*, 1998). These genes are members of the type II fatty acid synthase multi-enzyme complex (Cronan and Rock, 1996). This type of biosynthesis is widespread in bacteria but otherwise is restricted to the plastids of plants and algae. Within fatty acid biosynthesis, ACP (encoded by *acpP*) plays a

central role by holding the forming acyl chains. FabH and FabZ on the other hand are involved in the condensation (FabH) and dehydration (FabZ) steps of acetyl addition during acyl chain elongation. Apicoplast fatty acid biosynthesis has been implicated further by the discovery of another fatty acid synthase component in *P. falciparum*, *fabF*. This gene encodes β-ketoacyl-ACP synthase II (Cronan and Rock, 1996).

Lipid synthesis within the apicoplast seems probable following the identification of these nuclear encoded, apicoplast-targeted genes (Waller *et al.*, 1998). However, it could merely be a housekeeping function for the organelle. Conversely, it could be a compartmentalized site of fatty acid biosynthesis (Waller *et al.*, 1997). Hopkins *et al.* (1999), also suggested that the apicomplexan plastid could be the membrane lipid "factory" of the parasite. It was hypothesized that the lipids synthesized in the apicoplast were budded off in vesicular form, between the inner and outer plastid membranes and transported to adjacent organelles via extensions of its surface.

1.5.2.3 Amino acid biosynthesis?

It has been speculated that the apicoplast harbours a shikimate pathway for essential amino acid biosynthesis (Roberts *et al.*, 1998). The shikimate pathway is essential in higher plants, bacteria and fungi, but is absent from mammals (Kishore and Shah, 1988). Roberts *et al.* (1998) found that *in vitro* growth of *T. gondii*, *P. falciparum* and *Cryptosporidium parvum* was inhibited by the herbicide glyphospate, which inhibits the shikimate pathway enzyme 5-endopyruvyl shikimate 3-phosphate synthase (Kishore and Shah, 1988). In *T. gondii* and *P. falciparum* the inhibition of growth following treatment with this herbicide was reversed by the addition of *p*-aminobenzoate. This suggested the shikimate pathway supplies folate precursors for parasite growth. In the shikimate pathway, erythrose 4-phosphate and phosphophenol pyruvate are converted to chorismate in seven enzymatic steps. Chorismate is essential for the synthesis of *p*-aminobenzoate and folate and is also required for the synthesis of ubiquinone and aromatic amino acids.

When a partial chorismate synthase sequence was identified from the *T. gondii* sequence database it had an amino acid identity of between 44 and 51% with other chorismate synthases. However, no N-terminal extensions were seen on the *T. gondii* or *P. falciparum* homologues (Roberts *et al.*, 1998) suggesting the product has a cytoplasmic location.

As the genomic databases for apicomplexan parasites increase, other nuclear encoded genes whose products are targeted to the plastid will be identified which could indicate still further functions.

1.5.2.4 Plastid encoded open reading frames

Sequence analysis of the plastid genome of *Plasmodium* sp. has not indicated a cellular role for the extrachromosomal plDNA. As mentioned previously, most of its genes have been linked to gene expression (Wilson *et al.*, 1996). This system might be required for expression of one of the open reading frames (ORFs) of unknown function also present on the circle (Fig. 1.2). Mostly these are small, and it has been suggested they may code for additional very divergent ribosomal proteins (Wilson *et al.*, 1996). In *P. falciparum*, on the other hand, ORF470 codes for a large, reasonably well conserved protein (45-50% identity with other orthologues at the amino acid level) corresponding to the chloroplast gene *ycf*24 (Wilson and Williamson, 1997).

1.6 Open reading frame ORF 470 (*ycf*24)

1.6.1 Distribution

ycf 24, specifies the last major gene to be designated a function on the vestigial plastid genome of *P. falciparum*. This ORF (1414bp; 471 amino acids) is encoded by other apicomplexan pathogens, for example *Toxoplasma* and *Eimeria* (Denny *et al.*, 1998) and lies downstream of an LSU rRNA gene (Williamson *et al.*, 1994), (Fig. 1.2).

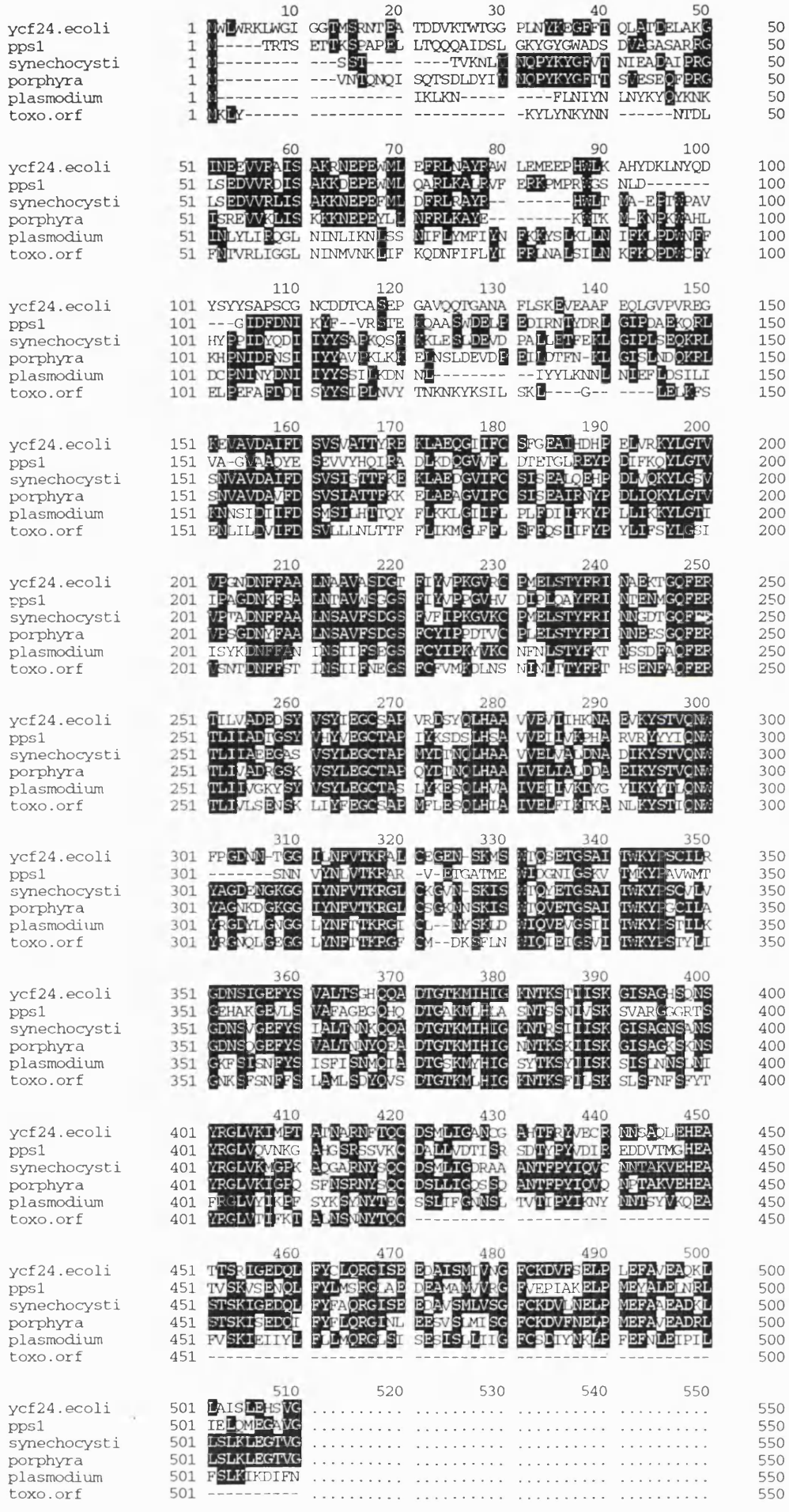
Although *ycf*24 has a limited distribution in known plastids, orthologues (45-55% amino acid identity) are found in a diverse array of organisms. Some examples are shown in the sequence alignment in Figure 1.9. This shows the predicted amino acid sequences of plastid encoded genes from two apicomplexan parasites (*P. falciparum* and *T. gondii*) alongside those from the red alga, *P. purpurea*, the Gram negative bacterium, *E. coli*, the cyanobacterium, *Synechocystis* PCC6803, and *Mycobacterium leprae*. The most marked conservation is observed towards the C-terminus.

*ycf*24 is also found in the plastid genomes of two diatoms, *Odontella sinensis* and *Skeletonema costatum*, the cryptomonad *Guillardia theta* and other red algae including *Cyanophora paradoxa* and *Cyanidium caldarium*. Furthermore, it is widely distributed on the nuclear genome of prokaryotes. These include the archaebacterium

Figure 1.9

An alignment of the hypothetical translation products of *ycf*24.

Shown are the predicted amino acid sequences of plastid encoded genes from two apicomplexan parasites [*Plasmodium falciparum*, *Toxplasma gondii* (partial)] and the red alga, *Porphyra purpurea*. Also included are sequences from the Gram negative bacterium (*E. coli*), the cyanobacterium (*Synechocystis* PCC6803) and *Mycobacterium leprae* (pps1).



Pyrococcus horikoshii, the eubacterium *Bacillus subtilis*, the cyanobacterium *Anabaena*, and *Mycobacterium tuberculosis* as well as those mentioned above. Conrad Mullineaux (UCL, personal communication) suggests that *ycf* 24 is likely to be common to all prokaryotes, indicating it has some functional importance.

It is interesting to note that although *ycf* 24 is absent from the plastid genomes of the green lineage, sequenced so far, it has been found on the nuclear genome of the higher plant *Arabidopsis thaliana*, where it has a putative chloroplast-directing signal sequence.

1.6.2 Unusual features of the *ycf* 24 orthologues

In two coccidian parasites, *Toxoplasma* and *Eimeria*, *ycf* 24 (465 and 387 amino acids respectively) encodes in-frame TGA stop codons. These occur at position 257 in *Toxoplasma* and at position 311 in *Eimeria*. Both putative stop codons are associated with a conserved tryptophan (TGG) codon (Denny *et al.*, 1998). Furthermore, in both *Toxoplasma* and *Eimeria*, the sequence flanking the unusual codon is highly conserved and interpretation of the opal stop as a tryptophan codon restores conserved C-terminal sequences. The presence of these stop codons has been suggested to be the result of codon capture: where the TGG codon was moved by directional A + T pressure to TGA. In other words, it is possible that stop codon suppression allowing full translation of the gene may occur, as has been recorded in other systems. Suppresser mechanisms include mutations in the peptidyl-transferase region of the large subunit rRNA (Jemiolo *et al.*, 1995) and specific deletions in the small subunit rRNA (Murgola *et al.*, 1988). Likewise, various mutations in tryptophan tRNAs can act as suppressers of UGA stop codons (Hirsh, 1971; Hirsh and Gold, 1971), as well as mutations in release factors and ribosomal proteins (Denny *et al.*, 1998). In apicomplexans it seems likely that a mutation in the SSU rRNA, which normally forms a duplex with UGA codons in the mRNA, is responsible for either partial or full suppression of the plastid UGA codons (Lang-Unnasch and Aiello., 1999). Of nine apicomplexan species examined, only the *Hepatozoan catesbia* plastid had the canonical 5′-UCAUCA-3′ SSU rRNA sequence responsible for UGA termination. All the others had the sequence 5′-UCUUUA-3′.

The *ycf* 24 orthologue of *Mycobacterium* sp. encodes a putative protein of 453 amino acids which shares 53% similarity with ORF470 from *P. falciparum*. An

interesting feature of this gene is the presence of an intervening sequence (intein) within the central region of the gene which is thought to be self excised (Davis *et al.*, 1992). The intein encodes a site-specific DNA endonuclease which is used for insertion of the intein coding sequence into other coding sequences (Gimble and Thorner, 1992; Perler *et al.*, 1992). In this sense, inteins may be regarded as the protein equivalent of the self-splicing, group I RNA introns. During the process of intein excision the peptide C- and N-terminal regions (exteins) are spliced together via transpeptidation to form a second functional polypeptide.

Inteins are not a common feature of proteins, few having been found in nature to date. Most of them occur in three important classes of conserved proteins. These include the vacuolar proton pump ATPase subunit of the yeast *Saccharomyces cerevisiae* (Kane *et al.*, 1990), DNA polymerase from *Thermococcus littoralis* (Hodges *et al.*, 1992) and the RecA proteins of *M. tuberculosis* (Davis *et al.*, 1992) and *M. leprae* (Davis *et al.*, 1994). Inteins have also been identified in *Synechocystis* including the DnaX protein of *Synechocystis* strain PCC6803, which is a homologue of the intein-less 71kDa tau-subunit of the *E. coli* DNA polymerase III (Lui and Hu, 1997). Although no phylogenetic relationship is seen between intein containing organisms, the presence of these inteins is thought to indicate a functional significance for the interrupted gene (Colston and Davis, 1994).

1.6.3 Function of *ycf* 24

The *ycf* 24 gene has not been designated a function, however, there is some evidence to suggest it is involved in metabolism. It has also been annotated as a subunit of an ATP-dependent transport protein (ABC transporter).

1.6.3.1 Metabolic role?

A possible metabolic function for the gene product of *ycf* 24 was suggested by M. Reith (Halifax, Nova Scotia, personal communication). Transcription of *ycf* 24 (ORF478) in the red alga *P. purpurea* was sensitive to light conditions. The level of mRNA transcripts was low in cells grown without light for three days, yet became abundant upon exposure to light. This observation contrasted with RNA levels of photosynthetic genes, where the transcript level remained high after three days without light. Biosynthetic genes exhibit similar variations in the level of transcription (Apt and Grossman, 1993). Reith's observation has not been followed up.

1.6.3.2 ABC transporter?

*ycf*24 from the cyanobacterium *Synechocystis* is annotated in the database as a gene for an ABC transporter subunit. However, we are not aware of any evidence to support this. ABC transporters are a large and diverse superfamily of proteins which mediate ATP-dependent movement of solutes across biological membranes (Higgins, 1995; Quentin *et al.*, 1999). Many ABC proteins are of considerable medical significance. An example is Pgh1, which plays a role in the malaria parasite's resistance to chloroquine. All ABC transporters share a common domain organization with four "core" domains that may be expressed as separate polypeptides (such as that of the oligo permease of *S. typhimurium*), or fused together in a number of configurations, including larger multi-domain proteins. The four core domains consist of two transmembrane domains which span the membrane multiple times forming the pathway through which the solute crosses the membrane. These domains determine the substrate specificity/selectivity of the transporter. The rest of the transporter is comprised of two ATP-binding domains. These are located at the cytosolic face of the membrane and couple ATP hydrolysis to solute movement. These latter domains share 30-40% sequence identity to all other ATP-binding domains from other ABC transporters, irrespective of the substrate specificity or species of origin. As a consequence, ATP-binding cassettes are easily distinguished (Higgins, 1995). The predicted amino acid sequence of *ycf*24 shows only a low level of similarity to ABC transporters. This is mainly confined to the extreme C-terminal domain which comprises four alpha helices (*E. coli* position ~455 onwards). Also a motif found in type 3-1 ABC transporters (Tomii and Kanehisa, 1998) commencing at *E. coli* position 198 is shared by *ycf*24. However, unlike other 3-1 group ABC transporter proteins, there are no Walker ATP binding motifs or putative membrane spanning protein sequences (Jose Saldanha, NIMR, personal communication).

On the other hand, a more typical ABC transporter subunit gene (*ycf*16) often lies adjacent to *ycf* 24 in bacterial genomes where ABC transporters tend to be encoded by genes within an operon (Quentin *et al.*, 1999).

1.6.4 *ycf*16

ycf 16 is a highly conserved open reading frame found on many chloroplast genomes, and is homologous to a gene of unknown function in bacteria. The predicted amino acid sequence suggests it is an ABC transporter (Wittpoth *et al.*, 1996).

Although absent from the plastid genome of apicomplexans, *ycf* 16 lies immediately downstream of *ycf* 24 in the cyanobacterium *Synechocystis*, as it does in the cyanelle genome of *Cyanophora paradoxa* and the most recently sequenced red algal plastid genome *Guillardia theta*. Like *ycf* 24, *ycf* 16 is absent from the plastid genomes of green lineage. However open reading frame homologies to *ycf* 16 are present in the nuclear genome of some plants currently being sequenced, such as *Glycine max* and *Lycopersicon esculentum* (K. Ellis, NIMR, personal communication). An association of *ycf* 24 with *ycf* 16 within these genomes has yet to be confirmed.

In all the prokaryotic genomes examined so far, *ycf* 24 and *ycf* 16 are in close proximity to one another. Usually they are adjacent, as seen in *Synechocystis* and *E. coli*, however, sometimes one or more genes separate them. This is observed in *Bacillus subtilis* and *M. leprae/tuberculosis*. The location of *ycf* 24 and *ycf* 16 in a variety of prokaryotes, protists and plants is given in Appendix 5 (courtesy of K. Ellis, NIMR). The close association of *ycf* 16 with *ycf* 24 in the plastid genomes of red algae and some diatoms implies that a *ycf* 16 orthologue might be nucleus-encoded in apicomplexans.

1.6.5 Co-transcription of *ycf* 24 and *ycf* 16: the *ycf* 24 operon

The *M. leprae ycf* 24 (ORF2) and *ycf* 16 (ORF4) are believed to be co-transcribed (J. Colston, NIMR, personal communication). This suggests the genes occur within an operon, i.e. are under the control of the same operator. Operons tend to contain groups of genes, which are associated with the same biosynthetic pathway. Therefore, if the function of one or more of the genes within an operon is known it may help to determine the role of genes of unknown function within the same operon. In *Mycobacterium* sp., evidence suggests that both the *ycf* 24 and *ycf* 16 genes occur within an operon comprising at least six open reading frames. Although no transcription has been confirmed, the putative genes contain overlapping nucleotides, which suggests they are transcribed together (J. Colston, NIMR). ORFs 2 and 4, correspond to *ycf* 24 and *ycf* 16 as mentioned above. ORF3 shows homology to *ycf* 24 but contains no intein. ORFs 5 and 6 on the other hand show similarity to *nif* genes (*nifS* and *nifU* respectively) (E. Davis and D. Jenner, NIMR, personal communication).

1.6.6 *nif* genes

The association of *ycf* 24 and *ycf* 16 in an operon containing *nif*, in particular *nifS* genes, is observed throughout prokaryotic genomes.

Nitrogen fixation is a property of certain bacteria and cyanobacteria, some of the former requiring a symbiotic relationship with leguminous plants. In the complex biochemical pathways of nitrogen fixation, N_2 is reduced to ammonia (NH_3), which is subsequently converted to an organic form. The reduction process is catalysed by the nitrogenase enzyme complex, which consists of two separate components: dinitrogenase and dinitrogenase reductase. At the genetic level many genes are involved in this process including the *nif* genes.

nifS encodes a protein required for the formation of the iron-sulphur core of the nitrogenase components. It is thought to be required for the desulphurization of L-cysteine to yield L-alanine and elemental sulphur, ultimately providing the sulphide present in the metalloclusters of the nitrogenase component proteins (Zheng *et al.*, 1994). Its presence in non-nitrogen-fixing organisms suggests a requirement for iron-sulphur complexes in other biochemical pathways.

1.6.7 Is *ycf* 24 an essential plastid gene?

The vestigial plastid genome of apicomplexans resembles that of the parasitic higher plant *E. virginiana*, yet the latter does not carry *ycf* 24. This counters the *raison d'être* suggested for maintenance of both these vestigial plastid genomes, i.e. dependence on the small number of open reading frames carried beside the otherwise similar and highly selected set of genes concerned with these organelles' expression systems (Wolfe *et al.*, 1992). One option for assessing whether retention of *ycf* 24 by the vestigial apicomplexan plastid is associated with an essential cellular process, rather than simple chance escape from deletion, would be to disrupt it using standard reverse genetics methodology. However, as will be discussed in this thesis, rather than study the gene in *P. falciparum* which lacks an organelle transfection system, two surrogate systems were chosen for practical reasons. These are the bacterial homologue in *E. coli*, and also that in the cyanobacterium, *Synechocystis* PCC6803. The latter is an organism that has been used principally for the study of oxygenic photosynthesis (Williams, 1988) and is an ideal tool for *in vitro* genetics as discussed below.

1.8 Syn*echocystis* PCC6803

1.8.1 Cyanobacteria

Cyanobacteria are the simplest organisms capable of carrying out oxygen-evolving photosynthesis. They are prokaryotes with an overall cell structure similar to that of Gram negative bacteria. Taxonomically cyanobacteria are divided into five sections. Sections I and II comprise the unicellular species which can occur singly or as colonial aggregates, held together by cell wall layers. By contrast, section III species are filamentous, forming a trichome (chain of cells) which grows by intercalary cell division. *Synechocystis* sp. belong to section I, which includes unicellular cyanobacteria that reproduce by binary fission or budding. However, *Synechocystis* sp. differs from the others in section I, as it is spherical and divides alternately in 2 planes at right angles to one another (Rippka *et al.*, 1979).

Unlike green or purple bacteria, cyanobacteria possess a second photosystem (PSII), allowing the removal of protons from water and resulting in the liberation of oxygen. They can also assimilate and metabolise inorganic nitrogen like higher plants, and some can fix atmospheric nitrogen (Porter, 1986). With the exception of the phycobilisome, the multi-protein complexes of the cyanobacterial photosynthetic apparatus are the equivalent to those of higher plants (Bryant, 1987). The role of phycobilisome in cyanobacteria is assumed in higher plants by chlorophyll *a/b*-binding proteins known as the light-harvesting complex II, for photosystem II, and the light-harvesting complex I, for photosystem I, (Thornber, 1986). The four other main complexes involved in photosynthesis are the PSII reaction centre and oxygen evolution complex, plastoquinol-plastocyanin reductase, the PSII reaction centre and ATP synthase. In addition to these membrane-associated complexes, many of the enzymes found in cyanobacteria (e.g. ribulose 1,5-bisphosphate carboxylase (Miziorko and Lorimer, 1983) and water soluble electron transport proteins (e.g. ferredoxin-NADP+ reductase; Yao *et al.*, 1984) are functionally and structurally similar to those in higher plants. It is not surprising, therefore, that the rapid growth rates of some cyanobacteria coupled with their simple structure and ease of handling make them an increasingly popular model system for studying oxygenic photosynthesis and inorganic nitrogen assimilation. There is however another significant difference from the photosynthetic apparatus of algae and higher plants that makes them particularly valuable: their amenability to genetic manipulation (Bryant and Tandeau de Marsac, 1988).

1.8.2 Natural competence to exogenous DNA

Cyanobacteria can be transformed by exogenously added DNA (Porter, 1988). They can also be the recipients of plasmid DNAs mobilised for transfer from *E. coli* (Wolk, 1984). The latter has been used to show that filamentous cyanobacteria are capable of heterocyst-based nitrogen fixation. However, I shall only be concerned in this thesis with natural competence for DNA uptake.

DNA mediated transformation is the most widely used method of genetic manipulation of cyanobacteria, first being described in the blue-green alga *Anacystis nidulans*, or *Synechococcus* PCC7943 (Shestakov and Khuyen, 1970). Since then, more cyanobacterial species have been found to have a natural ability to take up and assimilate homologous chromosomal DNA into their genomes. These include *Gloeocapsa alpicola*, or *Synechocystis* PCC6308 (Devilly and Houghton, 1977), *Agmenellum quadrupliatum*, or *Synechococcus* PCC73109 (Stevens and Porter, 1980), and *Synechococcus* R2 (Williams and Szalay, 1983). These are all species of *Synechococcus* and *Synechocystis* with an intermediate mol % G + C.

1.8.3 Transformation of *Synechocystis* PCC6803

The *Synechocystis* strain isolate, PCC6803, is a widely used study organism for photosynthesis as it has the ability to grow photoheterotrophically on glucose. This allows the propagation of mutants of photosystem II which are incapable of photosynthesis (Williams, 1988). It was first added to the list of cyanobacteria capable of high frequency chromosomal transformation by Grigorieva and Strestakov in 1982. They found that WT *Synechocystis* PCC6803 was competent, throughout exponential growth, for the uptake of DNA isolated from mutants of *Synechocystis* strains resistant to erythromycin. This property resulted in stable antibiotic resistant transformants after subculturing on non-selective media.

As mentioned above, an orthologue of the malarial plastid gene *ycf* 24 has been recognised on the genome of *Synechocystis* PCC6803 (Kaneko *et al.*, 1996; Tabata *et al.*, Cyanobase, DNA database, Japan), where it encodes 480 amino acids and is called *slr0074*. *Synechocystis* PCC6803, with its naturally occurring transformation potential and efficient transformation ability resulting in homologous recombination within its multi-copy chromosomes (Labarre *et al.*, 1989), was chosen as a surrogate system for studying the apicomplexan version of *ycf* 24. With approximately 10 copies of the chromosome per cell in culture conditions, cells may

be homoplasmic or heteroplasmic (i.e. containing both wild type and recombinant chromosomes). Thus, lethal genes can be recombined into the *Synechocystis* genome and selected for.

Insertional mutagenesis, by cloning antibiotic resistance genes into the genome of *Synechocystis* strain PCC6803, has been widely reported. Labarre *et al.*, (1989) transformed WT cells with either a pool of random *Hae*II DNA fragments of the host genome, which had been ligated with a chloramphenicol resistance gene (Cm^r), or *Ava*II kanamycin resistance marker ligated to random *Ava*II host DNA (Km^r) fragments. Integration of the resistance markers into the host genome resulted in stable Cm^r and Km^r transformants. These had no detectable phenotype on standard minimal media and had growth rates, in liquid culture comparable to the rate of WT. Southern blots, using the Cm^r, Km^r plasmid pLF8 as a probe, showed hybridization only to DNA isolated from the antibiotic resistant transformants not to wild type. In addition to this, an *Hae*II digest of the Cm^r transformants yielded a band corresponding to the size of the resistance gene, as did *Ava*II digestion of the Km^r transformants. This showed that integration of the antibiotic resistance genes restored its two flanking restriction sites, establishing that the resistance gene must be ligated at both ends to cyanobacterial DNA for insertion into the host genome. Homologous recombination between two *Hae*II sites (and thus two *Ava*II sites also) of the host genome was confirmed by the Cm^r transformant HC3. Here, insertion of the marker generated a small, but precise deletion between two *Hae*II sites originally present in the host genome. Homologous recombination was concluded to be occurring by gene conversion following a heteroduplex intermediate rather than by double crossover (Labarre *et al.*, 1989). This assumption was made mainly because of the mode of homologous recombination observed in other transformable bacteria (Stewart and Carlson, 1986). However, double crossover also seemed unlikely as single crossover events, resulting in the integration of the whole circular genome to yield a tandem duplication of the target genomic fragment spanning the resistance marker, was not observed.

Insertional inactivation on genes with a specific cellular function also has been carried out in *Synechocystis* strain PCC6803. For example Chitnis *et al.* (1989) used a kanamycin resistance gene to disrupt *psaD*. This gene encodes subunit II of photosystem I, which catalyses the photoreduction of ferredoxin, with plastocyanin as the electron donor. Southern blot analysis showed that all copies of the WT gene had

been lost in Km^r mutants selected for a few generations. Knockout clones had reduced photo-oxidation activity. Also, western blots on proteins isolated from the thylakoid membranes showed no signal for subunit II. Two years later the same was done for another gene encoding subunit III of photosystem I (*psaF*). Again homoplasmic mutants were obtained. This time, the photoautotrophic growth of the mutants was comparable with that of the WT, indicating that subunit III was dispensable for photosynthesis in *Synechocystis* species PCC6803 (Chitnis *et al.*, 1991).

In both of the PSI knockout experiments, homologous recombinants resulted after only a few generations of single colony selection. Segregation of single mutant clones by streaking onto fresh selection medium was found to be essential in mutant generation (Williams, 1988). This suggests that segregation of a disrupted non-essential gene occurs quite quickly as long as selective pressure is maintained.

1.9 *Escherichia coli*

E. coli was chosen as a second surrogate organism; having only one copy of *ycf* 24 per cell, knockouts should be generated immediately upon selection if *ycf* 24 was not essential. On the other hand, if *ycf* 24 is essential, disruption of the only endogenous copy would be lethal and no knockouts would be seen.

Homologous recombination in *E. coli* is dependent upon a functional *recA* gene. The protein expressed from this gene (RecA) is responsible for reannealing complementary DNA strands (Weinstock *et al.*, 1979) and the ATP-dependent pairing of superhelical DNA with homologous single stranded fragments (Shibata *et al.*, 1979). However gene-targeting using "foreign" DNA, such as that produced by PCR, is susceptible to degradation by endogenous restriction enzymes. Because of this, recombinant proficient strains need to be chosen which contain a functional *recA* gene but lack these restriction systems. For this reason, *E. coli* strain LE392 was used for knockout at the *ycf* 24 locus; LE392 (see materials and methods for the genotype) lacks the *E. coli* K restriction system but is *recA* positive making it a permissive host. In addition to this, it methylates incoming DNA (m_k^+) - this allows any DNA prepared from this strain to be used to transform any *E. coli* strain which contains the *E. coli* K restriction system.

1.9.1 Replacement using linear DNA

Gene targeting using linear double stranded DNA fragments in WT *E. coli* is generally inefficient due to the exonucleic activity of RecBCD. Although RecA carries out the fundamental pairing steps of recombination, other proteins are involved in other steps of the process. Paramount is the RecBC enzyme (encoded by *recB* and *recC*). This enzyme is important in DNA unwinding, its nuclease activities providing DNA with a free end, possibly allowing RecA to bind and begin recombination. Because of this exonucleic activity, recombination proficient strains in which the exonucleic activity of RecBCD is inactivated are used (e.g. Russell *et al.*, 1989; Shevell *et al.*, 1988). However, mutation of either *recB* or *recC* reduces recombination by approximately 100-fold, and this affects DNA repair. More recently an approach has been developed to overcome this difficulty by introducing Chi sequences (5′-GCTGGTGG-3′) at both ends flanking the insertion (Dabert and Smith, 1997). These sequences attenuate RecBCD exonuclease activities and stimulate its recombination activity (Dixon and Kowalczykowski, 1993). However, in 1999, Karoui *et al.* published another method describing how gene replacement could be achieved in many WT *E. coli* strains following electroporation of linear DNA. For chromosomal replacement they targeted the *E. coli* histidine (*his*) operon. A 3kb fragment of this operon was disrupted by a kanamycin resistance gene with or without Chi sites. Following linearization and purification, WT *his*$^+$ *E. coli* was transformed by electroporation. Kanamycin resistant colonies were patched on fresh LB-Km plates and replicated onto minimal medium lacking histidine as well as onto LB-Amp (ampicillin resistance being carried by the vector). Gene replacement events were scored as Km^r, His$^-$, Amp^s. They found approximately 60 gene replacement events/μg of linear DNA. However, the presence of Chi sites on the linear DNA had no effect on gene replacement. The presence of Chi sites only stimulated recombination if *E. coli* cells were made competent with $CaCl_2$ (Dabert and Smith, 1997). The success of recombination following electroporation suggested that this might partially inactivate RecBCD exonuclease activity, reducing degradation of linear DNA upon its entry to the cell, while still allowing the gene replacement event to occur.

Another method to stimulate gene replacement is to use λ recombination functions. This method is extremely efficient (Murphy, 1998). However, it requires

the use of particular *E. coli* strains, limiting its range whereas the method described by Karoui *et al.* (1999) is compatible with many *E. coli* strains.

1.9.2 Gene replacement using pKO3

Hamilton *et al.* (1989) first described a method for gene replacement in WT *E. coli* using homologous recombination between the bacterial chromosome and a plasmid whose replication activity was temperature sensitive. At the non-permissive temperature, antibiotic resistance could only be maintained if integration of the plasmid into the chromosome had occurred by homologous recombination between the cloned fragment and the bacterial chromosome. Subsequent switching of integrant clones to the permissive temperature allowed excision of the integrated plasmid. Depending upon the position of the second recombinational event resulting in excision, two genotypes were observed. The first retained the wild-type sequence with excision of the plasmid containing the altered sequence. The second resulted in replacement of the wild-type sequence with the altered sequence. There was however one major drawback: there was no selection for the loss of vector sequence.

The method described by Hamilton *et al.* (1989) was first modified by Bloomfield *et al.* (1991) to include a counter selectable marker in the chromosome, which would facilitate allelic exchange. They used the *sacB* gene from *B. subtilis* which encodes levansucrase, an enzyme that catalyses the hydrolysis of sucrose and levan elongation (Dedonder, 1966). When expressed in *E. coli* growing on media with sucrose, the *sacB* gene is lethal (Gay *et al.*, 1985). Link *et al.* (1997) decided to go one step further and combine the temperature sensitive origin of replication (*repA* derived from pSC101) with the negative selection marker (*sacB*) into one replacement vector, pKO3. They also included the *cat* gene (encoding resistance to chloramphenicol) on the vector as a positive marker for integration and distinguishing cells harbouring vector sequences following excision. Using this vector they introduced precise insertions and deletions, as well as insertion of a kanamycin resistance gene into both essential and non-essential *E. coli* genes.

1.9.2.1 Non-essential gene knockout (Link *et al.*, 1997)

The gene, *yjb*, encoding a highly abundant protein, of unknown function was disrupted using a kanamycin resistance gene and cloned into pKO3. Following transformation of a recombinant proficient *E. coli* strain at the permissive

temperature, 30°C, integrants were selected at the non-permissive temperature (43°C) by kanamycin and chloramphenicol resistance. Selected integrants were then serially diluted, at the permissive temperature onto plates supplemented with 5% sucrose and kanamycin. Viable colonies were classed as "potential knockouts" as kanamycin selected for the disrupted gene and growth on sucrose indicated excision and loss of pKO3. To confirm the loss of vector sequence, the potential positives were replica plated onto agar supplemented with 5% sucrose and chloramphenicol. Of these colonies, 98% were chloramphenicol sensitive. This indicated the loss of plasmid sequences and gene replacement, which was confirmed by PCR.

1.9.2.2 **Essential gene knockout** (Link *et al.*, 1997)

Two known essential genes *pepM* (a *map* gene encoding methionine aminopeptidase) and *rpsB* (a ribosomal protein encoding gene) were disrupted with a kanamycin resistance gene and cloned into pKO3. Following transformation at the permissive temperature, integration of pKO3 containing the disrupted gene was selected at the non–permissive temperature as above. Again potential integrants were serially diluted onto plates supplemented with 5% sucrose and kanamycin at 30°C. The positives were then replica plated onto 5% sucrose and chloramphenicol. This time all of the kanamycin resistant colonies remained chloramphenicol resistant. This indicated that plasmid sequences were still present in the cells and PCR confirmed that no knockout had occurred. Instead, excision of pKO3 containing the disrupted gene had occurred. Any colonies containing pKO3, with the disrupted gene, thereby allowing selection on both kanamycin and chloramphenicol in the presence of 5% sucrose, were attributed to either a mutation in *sacB* or a compensatory mutation in the genome rendering the expression of *sacB* non-lethal.

As will be described later, I chose this method to determine if disruption of WT *E. coli* at the *ycf*24 locus followed the pattern described by Link *et al.* (1997) for an essential or for an non-essential gene.

1.9 Summary

The discovery of the malarial plastid genome was unexpected and the source organism that contained the plastid of apicomplexans remains to be determined. Although descent from a dinoflagellate/ciliate clade was implied from earlier

phylogenetic studies of nucleus encoded rRNA genes, ciliates do not contain plastids. Unfortunately the high evolutionary rate of the plDNA has made it difficult to reach any clear conclusions on a red or green algal origin and this point is still debated. What seems clear however, is that the plastid originated by secondary endosymbiosis and the highly conserved gene content, and the order of gene sequences suggests the apicoplast evolved from a single photosynthetic progenitor. The high level of conservation, as well as evidence for transcription, translation, sensitivity to antibiotics and protein import all indicate a functional role for the apicoplast. What this is remains to be determined.

The plastid is of particular interest as a potential drug target and although research in this field is in its early stages, it aims to provide treatments effective against one of the most devastating diseases of mankind.

It is hoped that due to the plastid's prokaryotic nature, drugs affecting plastid biosynthetic pathways which do not affect those of eukaryotes may be developed. Initial evidence, although mainly *in vitro*, suggests that antibiotics can be targeted to the plastid resulting in inhibition of parasite growth. In addition to inhibition of organellar protein synthesis, drugs may also be developed which affect protein import into the plastid. This might involve the use of herbicides known specifically to inhibit the non-photosynthetic pathways of plastids. My thesis is concerned with establishing some indication of the function of the large open reading frame ORF 470 (*ycf* 24) encoded on the plastid genome. If this gene is essential even Ycf 24 may become a potential drug target.

Aims

I. This study aimed to "knock-out" *ycf*24 in bacterial surrogate systems thereby gaining insight into the function of the gene in *Plasmodium falciparum*. My first approach was gene inactivation in the cyanobacterium, *Synechocystis* PCC6803, replacing some copies of the multi-copy wild type gene with *ycf* 24 disrupted with either a kanamycin or a streptomycin/spectinomycin resistance cassette. Following transformation, recombinants were selected and their growth characteristics (phenotype) analysed with a view to determining the underlying effect of gene disruption. Similar experiments were planned using the *E. coli* single copy orthologue of *ycf* 24.

II. A second objective was to over-express *E. coli ycf* 24. As well as phenotypic observations, successful over-expression would allow purification of the protein for antibody production, which could give information on the location of the *ycf* 24 gene product and perhaps it's cellular function.

Chapter 2. Materials and methods

2.1 General buffers and solutions

The composition of general solutions and buffers is given in Appendix 1. General purpose reagents were supplied by BDH (British Drug Houses) and Sigma or NIMR General Supplies, unless otherwise stated.

2.2 Growth media

2.2.1 *E. coli* growth medium

LB* 10.00g Bacto-tryptone, 5.00g Bacto-yeast extract, 10.00g NaCl in 1 litre, pH7.0

SOC 20.00g Bacto-tryptone, 5.00g Bacto-yeast extract, 0.50g NaCl, 0.25mM $MgCl_2$, 20mM glucose, pH 7.0

2.2.2 *Synechocystis* growth medium

BG11* 10.00ml 100 x BG11[1], 1ml trace elements[2], 1ml iron stock[3], 1ml phosphate stock[4], 1ml carbonate stock[5] in 1 litre. Optional extras, 10ml TES buffer[6], 3g sodium thiosulphate.

*For LB or BG11 agar 15g of bactoagar (Difco) was added per litre.

Stock solutions:

[1]100x BG11	g/litre
$NaNO_3$	149.60
$MgSO_4.7H_2O$	7.49
$CaCl_2$	3.60
Citric acid	0.60
Na_2EDTA	1.12ml 0.25M solution, pH 8.0
[2]Trace elements	g/100ml
H_3BO_3	0.286
$MnCl_2.4H_2O$	0.181
$ZnSO_4.7H_2O$	0.022
$Na_2MoO_4.2H_2O$	0.039

$CuSO_4.5H_2O$	0.008
$Co(NO_3)2.6H_2O$	0.005
[3]Iron stock	g/100ml
Ferric citrate	0.60
or Ferric ammonium citrate	1.11

[4]Phosphate stock

3.05g in 100ml K_2HPO_4

[5]Carbonate source

2.00g in 100ml Na_2CO_3

[6]TES buffer

(N-tris(hydroxymethyl)methyl-2-aminosulfonic acid) 22.9g in 100ml, NaOH to pH8.2

2.3 Bacterial strains and growth conditions

Synechocystis strain isolate PCC6803 was grown photo-autotrophically in BG11 liquid medium (supplemented with 10mM sodium carbonate to promote growth) or on BG11 agar in a light incubator (10-20μE white light) at 30°C. For construct formation ligations were transformed in electrocompetent *E. coli* SURE® cells or XL1-blue MRF' (Stratagene).

E. coli strains were grown aerobically and usually at 37°C (temperatures used in conjunction with pKO3, are discussed later) in standard LB broth or LB agar. For cloning experiments transformation was by heat shock (42°C) into *E. coli* INVαF' (OneShot™) cells (InVitrogen), BL21(DE3)pLysS (Stratagene) and by electroporation into SURE® cells (Stratagene), recombination proficient strain LE392 (Stratagene), or Δ*recA* KM4104 [provided by Elaine Davis, NIMR (McEntee, 1977)]. The genotypes are given below.

Genotypes

INVαF' :	F', *endA*1, *recA*1, *hsdR*17 (r_k^-, m_k^+), *supE*44, *thi*-1, *gyrA*96, *relA*1, ϕ80*lacZ*ΔM15, Δ(*lacZYA-argF*)U169, λ-
BL21(DE3)pLysS :	F-, *ompT*, *hsdS*$_B$ (r_B^-, m_B^-), *dcm*, *gal*, λ(DE3), pLysS, Cm^r
SURE®:	e14$^-$(McrA$^-$), Δ(*mcrCB-hsdSMR-mrr*)171, *endA*1, *supE*44, *thi*-1, *gyrA*96, *relA*1, *lac*, *recB*, *recJ*, *sbcC*, *umuC::*Tn*5* (Kan^r), *uvrC*, [F', *proAB*, *lac*^{q}ZΔM15, Tn*10* (Tet^r)]

XL1-blue MRF': Δ(*mcrA*)183, Δ(*mcrCB-hsdSMR-mrr*)173, *endA*1, *supE*44, *thi*-1, *recA*1, *gryA*96, *relA*1, *lac*[F' *proAB* lac^{q}ZΔ*M*15 Tn*10* (Tet^{r})]^{c}

LE392: *hsdR*514 (r_k^-, m_k^+), *supE*44, *supF*58, *lacY*1 or Δ(*laclZY*)6, *galK*2, *galT*22, *metB*1, *trpR*55

KM4104: *mteA*, *strA*, *lysA*, *argA*, Δ$(lac)_{x74}$, Δ(*str-recA*), A234(*gal-bio*)

2.4 Analysis of growth of *Synechocystis* by colony area

After approximately two weeks growth on solid media the average colony area of wild type and inactivation mutants of *Synechocystis* was determined microscopically using the Optimas 5.2 programme according to the manufacturer's handbook.

2.5 DNA isolation

2.5.1 *Synechocystis*

Liquid cultures were supplied by Conrad Mullineaux and *Synechocystis* genomic DNA was isolated according to a protocol provided by Peter Nixon [Imperial College, London (Appendix 2)] or following the manufacturer's instructions for the Scotlab Nucleon Phytopure Plant Tissues DNA Extraction Kit. Both methods gave low yields, therefore, a modified version of the former method was used as follows. Cell pellets were subjected to three freeze-thaw cycles in liquid nitrogen. The cells were then resuspended in 400μl TES and incubated for 30 minutes with 100μl of 50mg/ml lysozyme at 37°C. After adding 50μl of 10% sarkosyl, the cells were vortexed in the presence of 600μl of phenol (Gallencamp Spinmix™) about once every minute for 15 minutes. After a 20 minute spin in a microcentrifuge at 13,000rpm, 600μl of phenol/chloroform was added and vortexed once a minute for 10 minutes. Finally 5μl of RNase (10mg/ml, Boehringer Mannheim) was added and the suspension incubated at 37°C for 30 minutes before continuing with the protocol described in Appendix 2.

2.5.2 *E. coli*

Total genomic DNA from *E. coli* was isolated using a slightly modified version of the standard method (Marmur, 1961). During this procedure a pellet of cells from the late logarithmic (or early stationary) phase of growth was resuspended in saline-EDTA buffer equal to $1/10^{th}$ - $1/40^{th}$ the volume of the original culture. SDS (BIO-

RAD) was then added to a final concentration of 1% and the cells lysed by gentle agitation in a 50-60°C water bath. Following lysis, signalled by a rapid increase in viscosity and a change in the cell suspension from turbid to opalescent, DNA was separated by phenol/chloroform extraction and ethanol precipitation (Sambrook *et al.*, 1989). If required, an additional RNA degradation step using DNase-free RNase (Boehringer Mannheim) was carried out for 1hr at 37°C prior to phenol/chloroform extraction.

2.6 Agarose gel electrophoresis (Aaij and Borst, 1972)

Fragments of DNA were separated by electrophoresis in 1% (w/v) agarose (Boehringer Mannheim or BIO-RAD) gels in 1 x TBE buffer containing 0.5μg/ml of ethidium bromide (Sigma). Samples were loaded after addition of agarose gel electrophoresis loading buffer (Sambrook *et al.*, 1989) and electrophoresis was carried out at 100V, constant voltage.

Nucleic acid markers

123bp ladder (GibcoBRL) at 300ng/μl with fragment sizes of : 0.123, 0.246, 0.369 (kb) etc.

Hind III-digested λ DNA (GibcoBRL) at 40ng/μl with fragment sizes of : 0.125, 0.564, 2.027, 2.322, 4.361, 6.557, 9.416, 23.130 (kb).

The molecular mass makers used in protein applications are given in section 2.25.

2.6 Estimation of DNA concentration

The concentration of DNA in solution was estimated by comparing the intensity of band fluorescence on an ethidium bromide-stained agarose gel with bands of known concentration from the λ*Hind* III molecular weight marker.

Primer concentrations were determined by UV spectrophotometry (see Section 2.9).

2.8 Restriction enzyme digestion of DNA

Various enzymes from (Boehringer Mannheim, GibcoBRL, Pharmacia, Stratagene and Promega) were used. Digestion was carried out usually for 2 hours at

37°C in the presence of the appropriate restriction buffer (supplied with each enzyme) unless otherwise stated by the manufacturer. Digestion of total genomic DNA from *Synechocystis* PCC6803 often required digestion for at least 16 hours in the presence of 100μg/ml BSA (New England Biolabs).

2.9 Primer design

Oligonucleotides were designed for the knockout of *Synechocystis ycf* 24 according to the genomic sequence provided by GenBank, Accession no. S76598.

Primers for the knockout and over expression of *E. coli ycf* 24 were designed from *E. coli* sequence provided by EMBL, Accession number D90811.

Primers were supplied by NIMR oligonucleotide synthesis service using an ABI 308B DNA synthesiser. They were supplied deprotected, at 0.1mM, in 35% ammonia and therefore required ethanol precipitation (Sambrook *et al.*, 1989). Following resuspension in dH_2O their concentrations were determined by UV spectrophotometry (1.0 at A_{260nm} = 33mg/ml for single stranded DNA) using a Beckman DU®-64 Spectrophotometer. Later, primers were supplied by Oswel at concentrations around 120-150μg/ml, dissolved in 1ml of sterile water.

Primer sequences and their uses are given in Appendix 3.

2.10 Amplification of DNA by PCR (Sakai *et al.*, 1988)

PCR was carried out in a volume of 50 or 100μl using PCR buffer II (Perkin Elmer) and 3mM $MgCl_2$ (Perkin Elmer). Each reaction contained 50-100ng of template DNA, 50pmoles of each of the two primers, 10nmoles each of dATP, dCTP, dGTP and dTTP (Pharmacia) and 1 unit of *Taq* polymerase (Perkin Elmer). Finally reactions were overlayed with approximately 50μl of mineral oil (Sigma) to prevent evaporation.

Direct PCR from recombinant *E. coli* colonies was carried out by resuspending cells from a colony of interest in 20μl of sterile water and heating at 95°C for 10 minutes to lyse the cells. Subsequently, the remaining reaction components were added: PCR buffer, $MgCl_2$, dNTPs, primers and finally *Taq*. PCR was then carried out as described below.

When extra proof reading was required, PCR reactions were carried out with 2.5-5 units of either *Pfu* or *PfuTurbo* DNA polymerase in place of Taq (Stratagene), using the buffers provided. DNA, primer and dNTP concentrations remained the same as described above.

PCR reactions were carried out on a Programmable Thermal Controller (MJ Research, Inc) and an Intelligent Heating Block (Hybaid Ltd.). Template DNA was initially denatured at 95°C for 30 seconds before undergoing a 1 minute annealing step at a temperature appropriate to the primers. Finally, an elongation step was carried out at 72°C (allowing 1 minute/kb for *Taq* and *PfuTurbo*, 2 minutes/kb for *Pfu*). All three steps were cycled for a total of 35 times.

Amplification of PCR products as large and larger than 10kb was carried out using LA*Taq* (TaKaRa). PCR using 2.5U of LA*Taq* polymerase was carried out in 50µl of LA*Taq* PCR buffer (TaKaRa) containing 2.5mM $MgCl_2$ (TaKaRa). As before, primer concentrations were 50 pmoles and dNTP concentrations remained at 10nmoles each. It was impossible to determine the DNA concentration as individual recombinant *E. coli* colonies were picked and used directly in the PCR. This time the DNA template was initially denatured by a 1 minute heating step at 95°C. This was followed by 30 cycles of 20 seconds at 98°C, 15 minutes at 68°C (annealing) and an extension step of 10 minutes at 72°C.

2.11 Isolation of plasmid DNA

Colonies were cultured overnight at 37°C with agitation in 3ml LB broth supplemented with the appropriate antibiotics. For minipreps, a minimum of 1.5ml of overnight culture was used. Plasmid DNA was isolated from colonies using the Wizard ® *Plus* Minipreps Purification System (Promega), employing a vacuum manifold according to the manufacturer's instructions. Plasmid DNA was eluted in 50µl dH_2O.

Alternatively (particularly for *endA* positive strains such as LE392 and BL21(DE3)pLysS and for automated sequencing) plasmids were isolated from 1.5ml of overnight culture using the S.N.A.P.™ (Simple Nucleic Acid Prep) miniprep kit from Invitrogen. Following this protocol, plasmid DNA was eluted in 60µl dH_2O.

2.12 Purification of the kanamycin resistance and *aadA* genes

A culture of *E. coli* containing the modified bluescript vector pBSSK coding for kanamycin resistance was provided by Mark Ashby at UCL. One colony was used to inoculate LB broth supplemented with 50μg/ml kanamycin monosulphate (Sigma) for plasmid preparation. The kanamycin cassette (Jiang *et al.*, 1987) was isolated by restriction with *Bam* H1 (Boehringer Mannheim), run into low melting temperature agarose and purified as mentioned below. To prevent self-ligation the kanamycin gene fragment was treated with the *E. coli* Klenow fragment of DNA polymerase I (Promega), resulting in a blunt ended fragment (Sambrook *et al.*, 1989).

The *E. coli aadA* gene cassette which encodes resistance to both streptomycin and spectinomycin (Goldschmidt-Clermont, 1991) was also provided by Mark Ashby. It was isolated from the recombinant bluescript vector (pBSHdSp1) by restriction with *Hin*d III (Boehringer Mannheim) and subjected to the same treatments as the kanamycin gene.

2.13 Cloning into plasmid vectors

2.13.1 Purification of insert DNA

PCR products were purified directly using the QIAquick PCR Purification Kit from QIAGEN according to the protocol provided. Digested PCR products and other DNA fragments were purified from agarose gels in a number of ways. The first method involved electrophoresis onto a DEAE cellulose membrane (Schleicher & Schuell), followed by elution in high salt elution buffer and phenol-chloroform extraction as described by Sambrook *et al.* (1989). A second method allowed the purification of DNA from low melting temperature agarose (May and Baker), by binding it to a silica matrix, using the Geneclean II kit (Bio 101, inc.). A third method (found to be the most efficient) involved the excision of the bands of interest from an agarose gel, and purifying the DNA using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's protocol.

2.13.2 Ligations

Ligations were carried out in 10μl reaction volumes using 50ng of vector and 1:1, 1:3 and 1:5 molar ratios of vector:insert DNA. Blunt end ligations tended to require a higher ratio of insert DNA than those with 3′ or 5′ protruding termini. For both

situations, however, incubation was at 14-17°C in T4 DNA ligase buffer (GibcoBRL) with one unit of T4 DNA ligase (GibcoBRL) overnight.

Occasionally PCR products were cloned directly into the plasmid cloning vector pCR® 2.1 using the Original TA Cloning® Kit from Invitrogen. This system utilized a vector containing 3′ dT-overhangs at its insertion site. As Taq polymerase tends to add a single dA to the 3′ ends of the PCR product in a non-specific manner (Clark, 1988) PCR products can be ligated without modification.

2.13.3 Transformation and selection of recombinants

2.13.3.1 *Synechocystis* PCC6803

Synechocystis cells were transformed directly according to the method described by Williams (1988). Liquid cultures of *Synechocysis* cells were grown to an optical density of A_{750} = ~0.5, diluted two fold and grown again overnight. The A_{750} (using a Phillips spectrophotometer) was then measured and the number of cells/ml calculated where cells/ml = A_{750} x 1.5 x 10^8. Samples of cells from 2-4mls of the overnight culture were harvested by centrifugation and washed once in fresh BG11 medium before resuspending in BG11 to 4 x 10^7 cells/ml. Aliquots of 100µl of these cells were then transformed by the addition of 1-5µg of DNA in TE buffer and incubated for 3-4 hours in white light at 30°C. To screen for primary transformants 20 and 50µl of the transformation mix were then spread onto two BG11 plates and allowed to grow without antibiotic selection. After three days recombinants were selected by overlaying with the appropriate antibiotic (kanamycin or streptomycin, Sigma) to a final concentration of 50µg/ml (in 5ml of top agar or in BG11 medium, 100µl). Agar plates were assumed to contain approximately 30ml of agar.

Every 10-14 days thereafter transformants were picked and re-plated on fresh BG11 plates supplemented with antibiotic to allow further segregation of any mutation. At the same time both fresh wild type (WT) plates were set up and transformants were plated onto BG11-only agar for growth comparisons in the absence of selection.

2.13.3.2 *E. coli*

E. coli host strains were transformed with 20-50ng of DNA by heat shock according to the manufacturer's instructions. This was usually at 42°C for 30 seconds,

with the exception of BL21(DE3)pLysS competent cells (Promega) where heat shock was for 45-50 seconds at 42°C. Usually 20, 50 and 100µl of each transformation was spread onto agar plates supplemented with the appropriate antibiotic/s and supplements (see results). In the case of the BL21(DE3)pLysS transformations (where the transformation efficiency was found to be low), cells were pelleted, resuspended in 200µl of SOC which was then spread onto the appropriate agar plates. INVαF′ cells did not require IPTG (Calbiochem®) to induce the *lac* promoter. Therefore X-Gal alone was required for blue/white colour selection of transformants. Accordingly, 40µl of 40mg/ml X-Gal (Melford Laboratories Ltd) in dimethylformamide was spread onto plates and allowed to diffuse into the agar for approximately 1hr, prior to plating cells.

Transformations of INVαF′ cells with the temperature sensitive, gene replacement vector pKO3 (kindly provided by Dr George Church, Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA) were carried out, and selected at 30°C, the permissive temperature for the temperature sensitive (*recA*) origin of replication.

Alternatively *E. coli* cells were transformed by electroporation using a BIO-RAD (Gene Pulser™) Electroporator, following the protocol provided by Stratagene. An electroporation cuvette with a 0.1cm gap (BIO-RAD) was used to obtain the maximum transformation efficiencies and cells were pulsed once for approximately 4-5ms (time constant) at 200Ω and 25µF. Some variation was required when applying a voltage across the 0.1cm gap: when using manufactured electrocompetent cells a voltage of 2.5kV was applied whereas "lab prepared" competent cells tended to have a higher salt concentration which meant that the voltage needed to be lowered to between 1.8 and 2.0kV.

"Lab prepared" electrocompetent cells were made by growing 1 litre of cells to an OD at A_{600} of between 0.5 and 0.7 (Beckman DU®-64 Spectrophotometer) before cooling on ice for 30 minutes. Cells were then harvested by centrifugation at 4,000rpm at 4°C for 10 minutes (Beckman, Model J-6B) and the pellet gently resuspended in 300ml of 10% ice-cold sterile glycerol. After a second centrifugation (as before) the pellet was resuspended in 100ml of 10% glycerol. Subsequent centrifugations resulted in the pellet being resuspended in smaller volumes of

glycerol, 50ml, 25 ml and finally 1ml. The cells were divided into aliquots of 120µl, flash frozen in liquid nitrogen and stored at -80°C until their use.

As for the heat shock transformations, 20, 50 and 100µl of each electroporation mix was spread onto agar plates containing the appropriate antibiotics (Sigma) and grown at 37° overnight. One exception were transformations using the pKO3, temperature sensitive, gene replacement vector (Dr George Church, Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA). Here transformed LE392 cells containing circular episomes were selected initially at 30°C, the permissive temperature for the M13 temperature sensitive origin of replication. Recombinants in which integration of the vector plus insert into the *E. coli* genome had occurred were selected next at the non permissive temperature, 42°C. Finally, potential "knockout" clones were re-selected at the permissive temperature (see results Chapter 3.3).

2.13.4 Analysis of recombinants

Recombinant clones were analysed for the presence of recombinant plasmids by restriction digestion (see Results) and sequencing (both manual and automated, see 2.14). Genomic recombination events were confirmed by Southern blotting (Section 2.15).

2.14 Sequencing (Sanger *et al.*, 1977)

2.14.1 Manual DNA sequencing

Manual sequencing of plasmid DNA was carried out by the dideoxy chain termination method (Sambrook *et al.*, 1989) using Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical) in accordance with the manufacturer's protocol, except that an initial heat denaturation step was included (Mass and Roop, 1992). Plasmid DNA (~1µg) and primer (~0.5pmol) were heated at 100°C for 5 minutes and then chilled on ice for 5 minutes. Sequenase reaction buffer was added (2µl) and the labelling and termination reactions carried out following the manufacturer's instructions. Compressions and "ghost bands" were resolved by the addition of deoxynucleotydyltransferase (USB), as described by Fawcett and Barlett (1990). Sequencing reactions were run on 6% denaturing polyacrylamide gels made with SequaGel™ ready-to-use gel casting solutions (National Diagnostics) and run in

1 x TBE buffer at 55W using an S2 sequencing gel apparatus (Gibco BRL), as described by Sambrook *et al.* (1989). Gels were fixed for at least 20 minutes in 10% v/v acetic acid and 20% v/v methanol, dried under vacuum at 80°C (Model 1583, BIO-RAD) and exposed to BioMAX-MS film (Kodak, supplied by Amersham) at room temperature for at least 16 hours.

2.14.2 Automated DNA sequencing

Automated sequencing was carried out using the ABI PRISM™ dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems), according to the protocol provided with the kit. Approximately 300ng of plasmid containing the DNA insert of interest were used in each sequencing reaction. Both external (designed against vector sequence) and internal primers (designed against the insert) were used in sequencing reactions. A list of primers and what they were used for is shown in Appendix 3). The *ABISEQ1 programme used for the sequencing reaction included an initial denaturation step of 95°C for 45 seconds, an annealing step for 30 seconds at a temperature appropriate for the primer and an elongation step of 60°C for 2 minutes. This was repeated for 25 cycles. Following the labelling reaction excess dye terminators were removed by ethanol precipitation, in accordance with the manufacturer's instructions, and loading buffer added immediately before loading samples on a gel (components from Perkin Elmer). Samples were separated on an ABI Prism™ 377 sequencer by Malcolm Strath (NIMR).

2.15 Southern analysis

2.15.1 Southern transfer

Southern blotting (Southern, 1975) was carried out as follows. After agarose gel electrophoresis, DNA was denatured in 0.2M HCl and neutralized in 0.5M NaOH, 1.5M NaCl by submerging the gel in each solution for 15 minutes at room temperature. Capillary transfer to Hybond N+ (Amersham) or GeneScreen Plus Hybridization Transfer Membrane (NEN™) in 0.5M NaOH, 1.5M NaCl was carried out as described in Sambrook *et al.* (1989). Following transfer the blot was washed in 2 x SSC and placed on 3MM paper (Whatman) and the DNA cross-linked by UV (autocrosslink, UV Stratalinker 1800, Stratagene). The blot was covered in Saranwrap (Dow) and stored at 4°C, or longer at -20°C, until used for hybridization.

2.15.2 Template preparation for hybridization

PCR products for *ycf* 24, or isolated and purified antibiotic resistance genes, were used as templates for probe synthesis. DNA was run into and excised from low melting point (LMP) agarose (May & Baker). The gel slice was then diluted with 3 volumes of dH_2O and melted at 65°C and stored at -20°C until used.

2.15.3 Labelling reaction

In the labelling reaction, the Prime-It® II Random Primer Labelling Kit (Stratagene) was used with approximately 25ng of DNA according to the protocol provided with the kit. The radioisotope used was α-[^{32}P]-dATP from Amersham.

2.15.4 Purification of the radiolabelled probe

Separation of unincorporated nucleotides from the probe was achieved according to the manufacturer's instructions by purification through NICK® Spin Columns (Pharmacia). TE was used as the equilibration buffer and the probe was eluted in 150µl of TE buffer.

2.15.5 Pre-hybridization and hybridization to the radiolabelled probe

Non-specific binding sites on the membrane were blocked by incubation at 65°C for at least one hour with 10-20ml (depending upon the size if the blot) of hybridization solution containing denatured salmon sperm DNA (Stratagene). This was carried out in a Micro-4 hybridization oven (Hybaid™). The purified probe was then denatured by boiling for 10 minutes, snap-cooled on ice for 10 minutes, and then added to the hybridization tube. The membrane was then incubated overnight rotating at 65°C, in a hybridization oven (Hybaid).

2.15.6 Post-hybridization

Unbound probe was removed from the membrane by washing twice for 10 minutes at room temperature with 250ml of Southern wash 1 (Appendix 1). This was followed by a 10 minute 65°C wash with 250ml of Southern wash 2 (Appendix 1). If the radioactive signal was still high a final wash for 10 minutes at 65°C was undertaken with 250ml of Southern wash 3 (Appendix 1). The blot was then wrapped

in Saran wrap (Dow) and was exposed to XB-200 High Definition X-Ray Film (X-ograph Imaging Systems) at 70°C for an appropriate time (usually overnight).

2.16 Stripping of Southern blots

All of the radiolabelled probe was removed from Southern blots by immersing the blot in boiling dH_2O/2%SDS and cooling to room temperature with constant shaking (Hot Shaker, Bellco®). This process was repeated 4 times after which the blot was exposed to film, as described above. If any signal remained the process was repeated, however, this was unusual.

2.17 Copy number analysis of inactivated *Synechocystis ycf*24

Doubling dilutions of a PCR product for *ycf* 24 (16ng-0.125ng) was used as a control and also as a probe in a Southern blot. The intensity of the signal was measured using a phosphorimager (Storm 860, Molecular Dynamics) and the appropriate software (Image Quant, Molecular Dynamics). The hybridization efficiency of the probe was calculated by comparison of the control results and the copy number determined from the results obtained using total DNA.

2.18 Transmission electron microscopy (TEM) (Stanier, 1988)

(Protocol carried out by Elizabert Hirst, NIMR)

2.18.1 Fixation

Samples of *Synechocystis* from either liquid culture or collected from plates and resuspended in BG11 medium were pelleted by centrifugation and prefixed by incubation for 30 minutes at room temperature with 0.5% glutaraldehyde in SCB (0.1M sodium cacodylate at pH 7.2). The cells were then fixed overnight with 3% glutaraldehyde in SCB.

2.18.2 Agarose embedding

2% of aqueous agarose (high melting point) was melted 60°C and cooled to 37°C in water bath. The fixed *Synechocystis* samples were then pelleted at 6,500rpm and most of the fixative aspirated. Following resuspension in 0.3ml agarose at 37°C, the samples were again pelleted in Eppendorf tubes. The agarose was set by placing

the tubes on ice with a little 3% glutaraldehyde carefully pipetted onto the surface. Finally, the samples were trimmed with a razor to $1mm^3$ for EM.

2.18.3 Refixation and staining

The specimens were initially refixed in 3% glutaraldehyde in SCB and washed 4 times with SCB. Post fixation, the samples were immersed in 1% OsO_4 in SCB pH7.2 at room temperature for 30 minutes and for 3 hours at 4°C. After washing with distilled water, samples were stained by immersion for 1½ hours in 1% aqueous uranyl acetetate.

2.18.4 Dehydration

The specimens were dehydrated by submersion at room temperature for 10 minutes each in 50%, 75% and 95% ethanol. This was followed by three 10 minute incubation steps with 100% EtOH.

Subsequently the samples were further dehydrated for 20 minutes in propylene oxide (PO) and in PO: araldite (2:1, 1:1 and 1:2) for 45 minutes each. Finally the samples were infiltrated with several changes of araldite (CY212, Agar) over 8 hours and polymerized at 70°C overnight.

2.18.5 Further staining

The specimens were stained in saturated ethanolic UA (uranyl acetate) followed by Reynolds lead citrate (Appendix 4).

The cell sections were viewed in a JEOL 100 CX transmission electron microscope.

2.19 Scanning electron microscopy (SEM)

(Carried out by Elizabeth Hirst, NIMR, according to Hayat, 1978)

2.19.1 Fixation

Samples of *Synechocystis* cells were washed in 1 x PBS and dried on cover slips coated with 2µl poly-L-lysine (Sigma) at a concentration of 100µg/ml. Following fixation at room temperature by incubation for 45 minutes in 0.1M SCB/2% gluteraldehyde, the slides were washed with 2% cacodylate buffer, pH7.2.

2.19.2 Post fixation

Post fixation, the slides were dehydrated in ethanol: 5 minutes with 50% EtOH, 10 minutes with 70% EtOH, 10 minutes with 95% EtOH and finally for 30 minutes with 100% EtOH. The slides were then immersed for 15 minutes in HMDS (heXamethyl disilazine) and allowed to air dry at room temperature. The cover slips were then mounted onto SEM stubs and desiccated overnight before being sputter coated with 20nM gold (Sputter Coater, EMscope).

The samples were viewed in a JEOL CF35 scanning electron microscope.

2.20 DAPI staining

Nucleoids were stained with DAPI (Sigma) at 0.01μg/ml, following a protocol similar to that described by Hiraga *et al.* (1989). 2μl of poly-L-lysine (Sigma), at a concentration of 100μg/ml, was spotted onto glass slides and allowed to dry. Cell samples of either *E. coli* or *Synechocystis* were washed twice in PBS and 2μl spotted onto the poly-L-lysine and air dried. The slides were then fixed with methanol and washed with dH_2O. 2μl of DAPI stain was placed onto the fixed cell samples and incubated at room temperature for approximately 20 seconds before being washed off with dH_2O. Localization of the nucleoids under Vectashield mounting medium (Vector Laboratories Inc.) was detected under UV light and analysed using conventional UV-light microscopy (Zeiss) and Ectachrome P1600 film.

2.21 Giemsa staining

Slides were prepared as described above using poly-L-lysine and fixed in methanol. The samples were stained with Giemsa's staining solution (BDH) diluted to 10% in water. Following air drying, samples were analysed using light microscopy (Zeiss) and photographed using P1600 film (Fuji).

2.22 Autofluorescence of *Synechocystis*

(Carried out by Iain Wilson, NIMR)

Slides were prepared as described above with washed WT and mutant *Synechocystis* cells. Following fixation, the samples were observed directly under UV light and pictures taken with Ectachrome P1600 film. Cells with normal amounts of chlorophyll showed bright orange fluorescence.

2.23 Statistical analysis of SEM data - The Chi-square test.

The statistical significance (*P* value) of the proportion of cells undergoing division in *Synechocystis ycf*24 mutants when compared to wild type (WT) cells also in division, was determined using the chi-square test (Croxden and Cowden, 1962). The χ^2 value was determined following the formula below:

$$\chi^2 = \sum \frac{(f - f_c)^2}{f_c}$$

Where: f = Observed frequency, i.e. the number of mutants undergoing division.

f_c = Computed frequency, i.e. the number of cells expected to be undergoing division, in this case the number of WT cell in division.

As only two criteria were analysed, either dividing or not-dividing, the *n* (degrees of freedom) value for the experiment was 1. This value, with that calculated for χ^2 were used to look up the specified value of *P* (Appendix J, Croxden and Cowden, 1962). *P* values <0.001 were assumed to be significant.

2.24 Analysis of *Synechocystis* cell cycle

(Carried out by Conrad Mullineaux, UCL)

Live *Synechocystis* cells growing on BG11 plates were monitored by a video linked phase contrast microscope (Zeiss Axiophot) with images taken over 5-6 generations to monitor cell cycle length. Images were recorded at 3 hour intervals for the first 12 hours and at time points 23hrs, 30hrs and 47hrs thereafter. The state of dividing cells within a colony was then analysed. In between time points, the plates were returned to the light incubator at 30°C (10-20μE light). Whole field images were analysed by A. Law at the cellular level.

2.25 PAGE

Samples were solubilized in SDS sample buffer by heating at 100°C for 5 minutes and analysed by SDS-PAGE on homogenous 10% SDS/polyacrylamide (w/v) gels according to the method of Laemmli, 1970.

Soluble proteins were analysed by electrophoretic separation in native (non-denaturing) polyacrylamide (12% w/v) gels using native loading buffer.

Electrophoresis was carried out at 25-40mA using a Mighty Small II vertical slab gel unit (Hoeffer Scientific Instruments) or using the Xcell II™ Mini-Cell (NOVEX™) according to the manufacturer's instructions. Pre-cast gels were used in conjunction with the Xcell II™Mini-Cell. For native applications, 10% Tris-Glycine gels were used with the supplied running buffer. For denaturing applications, 10% NuPAGE® Bis-Tris gels were used, with MOPS running buffer, (both supplied by NOVEX™).

Following electrophoresis, proteins were stained under agitation (Orbital Mixer, Denley) for 1 hour at room temperature with 0.1% (w/v) Coomassie Brilliant Blue R-250 (Sigma). This was followed by several incubations in destaining solution until the separated proteins and markers were easily visualized.

Molecular mass markers

Prestained Protein Marker, Broad Range (New England Biolabs) with marker sizes in kDa of: 6.5, 16.5, 25, 32.5, 47.5, 62, 83, and 175.

Unstained Wide Range Protein Standard Marker (NOVEX™) with marker sizes in kDa of : 2.5, 3.5, 6.0, 14.4, 21.5, 31.0, 36.5, 55.4, 66.3, 97.4, 116.3 and 200.

Unstained Protein marker, Low Range (Pharmacia) with marker sizes in kDa of: 20, 30, 45, 68 and 90.

2.26 Western blotting (Burnette, 1981)

2.26.1 Protein transfer to nitrocellulose membrane

Protein samples were run on a 10% SDS or native polyacrylamide gel and electrophoretically transferred to Hybond™-C Pure nitrocellulose membrane (Amersham) using the Sartoblot®II-S semi-dry blotting apparatus (Sartorius), according to the manufacturer's instructions. Transfer of the proteins was carried out at 25-50mA (depending on the size of the blot) for three hours.

If unstained protein marker was used, blots were subsequently stained with Ponceau-S stain (Sigma) and washed with dH_2O until the marker bands were easily

observed. These were marked with pencil and the remaining stain removed by further washing before continuing with the protocol.

2.26.2 Probing and developing of the membrane

Western blots to be probed with antibody were blocked with 5% (w/v) non-fat milk powder (Marvel) in 1 x PBS/0.1% Tween 20 (v/v) (Sigma) with agitation for 1 hour at room temperature. The membrane was then washed three times, for 10 minutes in PBS/0.1% Tween 20 (v/v), before incubating for 1 hour at room temperature (or overnight at 4°C) with agitation in solutions containing various primary antibodies. Those used were murine monoclonal anti-GST (Zymed) diluted 1/1,000 in PBS/Tween, rabbit polyclonal anti-MBP diluted 1/10,000 in PBS/Tween (New England Biolabs) and a mouse penta-His Antibody (QIAGEN) diluted 1/1,000 in PBS/ Tween. I also used anti-FtsZ kindly provided by Dr H. Erikson (Duke University, North Carolina, USA.), diluted 1/1,000 in PBS/Tween. Unbound primary antibody was removed by 3x 10 minute washes in PBS/Tween. This was followed by incubation with agitation for 1 hour at room temperature with various second antibodies conjugated to horseradish peroxidase. Rabbit anti-mouse IgG (Zymed) diluted 1 in 1,000 in PBS/Tween was used for anti-GST and anti-His probed western blots. Goat anti-rabbit IgG (BIO-RAD) diluted 1 in 3,000 in PBS/Tween was used for anti-MBP and antibody-probed western blots. The presence of peroxidase was detected by chemiluminescence using ECL™ Western Blotting Detection Reagents (Amersham) according to the manufacturer's instructions.

2.27 Immunofluorescence

(Carried out by Iain Wilson, NIMR)

Bacteria were washed twice in 1x PBS and dried on glass slides coated with poly-L-lysine (Sigma), at a concentration of 100μg/ml, and fixed with methanol (as described previously). Cells were then permabilized by incubation for 15 minutes at room temperature in 0.1% Triton X-100 with 1mg/ml of lyzozyme (Sigma). Localization of anti-MBP (New England Biolabs) was detected with an FITC-labelled anti-rabbit IgG (Sigma) using Vectashield mounting medium (Vector Laboratories Inc.) and observed under UV light.

Prolonged incubation periods of ~1.5 hours with anti-FtsZ (Dr H. Erikson, USA) were used to detect FtsZ as recommended by Yu and Margolin (1999) in *E. coli* cells over-expressing Ycf 24 as a fusion protein.

Fixed fluorescent specimens were analysed using an Olympus microscope with CCD camera operated by DeltaVision computer software (Applied Precision Inc., USA). Other images were recorded using a conventional UV-light microscope at x630 magnification and Ectachrome p1600 film.

2.28 Purification and cleavage of maltose binding protein (MBP)

Following cloning into pMAL-c2 (New England Biolabs) and induction with 0.3mM IPTG (Calbiochem), Ycf 14 was over-expressed as an MBP fusion protein. Recombinant fusion protein was purified under native conditions according to the Protein Fusion and Purification System instruction manual provided with the vector. Fusion proteins from 1ml of induced culture were selectively bound to 15ml of amylose resin (New England Biolabs) in a 2.5 x 10cm column (BIO-RAD) at a flow rate of approximately 1ml/minute. After washing, the fusion protein was eluted (at the same flow rate) with 10mM maltose (Sigma) and 20-40 fractions of 1.5ml were collected. Elution of the fusion protein was detected by UV absorbance at 280nm using liquid chromatography (Pharmacia).

Cleavage with factor Xa (New England Biolabs) was carried out according to the manufacturer's instructions. A ratio of 1% the amount of fusion protein was used (i.e. 1mg of factor Xa within a reaction containing 100mg of fusion protein). Incubation was at 4°C over a period of 3hrs to several days.

To partially denature the fusion protein and allow cleavage to occur, low concentrations of SDS (BIO-RAD) were added to the reactions, as suggested by the manufacturer, ranging from 0.005-0.5% (see Results).

2.29 Purification of polyhistidine-containing recombinant proteins

Following cloning into pET-28a (Novagen) and induction with 0.4mM IPTG (Calbiochem®) Ycf 24 was over-expressed as a His-tagged fusion protein. From an induced culture of BL21(DE3)pLysS cells expressing Ycf 24, 1ml was subjected to a protein miniprep under native conditions as described by protocol 13 in the QIAGEN handbook (High-level expression and purification of 6x His-tagged proteins, third

edition). Proteins bound to Ni-NTA agarose (QIAGEN) were eluted with 250mM imidazole.

Purification of recombinant fusion protein was also carried out using the Xpress™ System Protein Purification (version C) from Invitrogen according to the manufacturer's instructions for bacterial lysates. Proteins were bound to nickel resin columns (ProBond™) prepared either for purification under native or denaturing conditions. Under native conditions proteins were eluted by increasing concentrations of imidazole or at pH 4.0. Under denaturing conditions, proteins from cells previously lysed in the presence of 6M guanidinium hydrochloride, were eluted at pH4.0 in the presence of 8M urea.

Purification of inclusion bodies was carried out by Kaveri Rangachari (NIMR) according to the method described by Zhao *et al.*, (1999) for the purification and refolding of the *Streptococcus pneumoniae* penicillin-binding protein 2a, over-expressed in *E. coli.*

2.30 Over-expression and cleavage of glutathione-S-transferase (GST) fusion proteins

ycf 24, cloned into pGEX-6P-1 (Pharmacia) and induced with 0.3mM IPTG, was over expressed in 20ml of culture for 2 hours. Following centrifugation at 2,000rpm (Centaur 2, MSE) for 20 minutes the pellet was resuspended in 2ml of PreScission protease buffer (500mM Tris.HCl, 10mM EDTA, 150mM NaCl, 1M DTT at pH7.0). Cells were lysed by incubation for 30 minutes on ice with 5µl of 50mg/ml lysozyme (Sigma) and sonicated; 15 second pulses at 10 second intervals for a total of 2.5 minutes (VibraCell[a] VC600 sonicator from Sonics & Materials). Soluble protein was separated from cell debris and insoluble protein by ultra-centrifugation at 50,000g for 1 hour (Beckman L7-65 ultracentrifuge). Soluble proteins were concentrated 100x into the same buffer using Centricon® concentrators (Amicon®). 50µl of concentrated soluble proteins were incubated with 2.5U PreScission™ protease (Pharmacia) overnight at 4°C.

Chapter 3. Results

3.1 Knockout of *ycf*24 in *Synechocystis* PCC6803

3.1.1 Insertional inactivation of *ycf*24 in *Synechocystis*

A central segment (delineated by two *Hind* III sites approximately 1.0kb apart) of wild type *ycf* 24 from *Synechocystis* sp., strain PCC6803, was amplified from genomic DNA by PCR using the primers A3 and A5 (see Appendix 3). The amplified fragment was then digested with *Hind* III endonuclease and cloned into pUC9. Following transformation of *E. coli* SURE cells, recombinants were selected by growth on ampicillin and blue/white colour before being analysed by restriction with *Hind* III. The correct sequence of both 5′ and 3′ insert junctions for a number of recombinant clones was confirmed by manual sequencing using the primers P1 and P2.

For the disruption of *ycf* 24, a kanamycin resistance gene, Km^r (Jiang *et al.*, 1987), was isolated by restriction with *Bam* HI from a modified bluescript vector, pBSSK (courtesy of Mark Ashby, UCL) and blunt ended using the klenow fragment of DNA polymerase I. The Km^r gene was subsequently cloned into a unique Klenow-treated, *Xba* I enzyme site within the *ycf* 24 insert. Recombinants in XL1-blue MRF′ cells were selected on BG11 plates supplemented with 50μg/ml kanamycin. Disruption of *ycf* 24 in six of these recombinant clones was confirmed by sequencing the junction regions using the primers K1 and K2. Four of these clones (3, 4, 5 and 8) were used for transformation of *Synechocystis* pCC6803.

Transformation of *Synechocystis*, using selection with kanamycin after three days was successful in each case, yielding hundreds of primary transformants. After one week in the presence of kanamycin it was evident that the primary transformants were much slower growing than the wild type; colonies of the latter were visible within four days where as it took two weeks for the mutants to be seen with the naked eye. Under a light microscope the mutant colonies were small with a very ragged outline (Fig. 3.1.1A). In contrast those of the wild type were larger with smooth-contours (Fig. 3.1.1B).

It should be mentioned here that a *Synechocystis* pCC6803 clone, containing a kanamycin resistance gene randomly inserted within a non-coding region (Emlyn-Jones *et al.*, 1999), resulted in colonies with the wild type morphology when grown under the same conditions as the *ycf* 24 mutants.

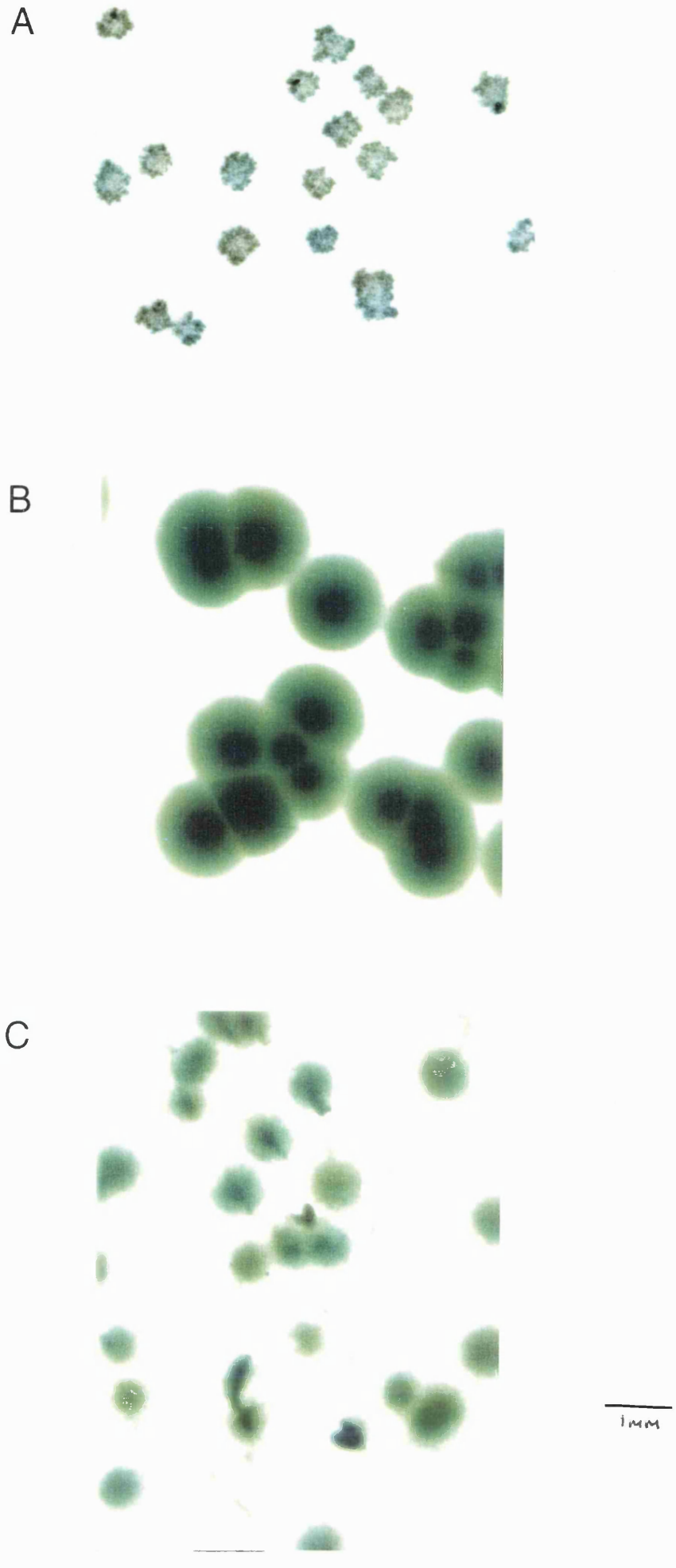
Figure 3.1.1

Colony morphology of *Synechocystis* PCC6803 transformants.
(All at the same magnification)

A. Primary kanamycin resistant transformants of *Synechocystis* following transformation with pUC9 containing disrupted *ycf* 24. The resulting colonies after two weeks growth on BG11 agar supplemented with 50μg/ml of kanamycin are small with a severely ragged colony morphology.

B. Wild type *Synechocystis* PCC6803 showing large colonies with a smooth colony morphology after two weeks growth on BG11 agar.

C. Kanamycin resistant transformants following the first round of selection. Colonies remain smaller than WT and although still ragged, they are not as ragged as the original transformants.


A
B
C
1MM

DNA was extracted from three "ragged" mutant clones (T3, T4, T6 and T8) and digested with *Hind* III. Southern blots were prepared by electrophoresis of 200ng of DNA isolated from each clone on an agarose gel and hybridization with probes for either wild type *ycf* 24 or the kanamycin resistance gene. The wild type *ycf* 24 probe hybridized to restriction fragments of three expected sizes; a 1kb band which corresponded to the wild type form of *ycf* 24 and bands of 1.3 and 0.5kb corresponding to the inactivated gene (Fig. 3.1.2). The Km^r probe hybridized to two restriction fragments only. The first of these was a band of 1.3kb previously shown up when *ycf* 24 was used as a probe and which corresponded to the inactivated gene. The second was a band of around 400bp, corresponding to an internal kanamycin fragment resulting from restriction at the two *Hind* III sites within Km^r (see Fig. 3.1.2 and Fig. 3.1.3). A very faint signal was seen around 500bp in some cases. However the significance of this is unclear, as the first *Hind* III site within the kanamycin resistance gene occurs close to the 5′ end and hence, the 500bp *ycf* 24 fragment was unlikely to hybridized to the kanamycin probe. As well as confirming that *ycf* 24 had been inactivated these data show that the primary transformants were heteroplasmic for the disrupted form of *ycf* 24.

As each cell of *Synechocystis* contains multiple copies (up to 10) of its genome, attempts were made to force further segregation of the heteroplasmic mutant genotype and generate a complete knockout (homoplasmy). To do this mutant colonies were picked every two weeks from each of the transformed clones and streaked onto fresh BG11 plates supplemented with kanamycin. Also, to determine if the cells would lose the inactivated gene and resume normal growth in the absence of selection, some mutant colonies were streaked on plates lacking kanamycin.

Following the first round of selection the kanamycin resistant mutants remained both slow growing and ragged in colony morphology, although not as severe as the initial transformants (Fig. 3.1.1C). Relaxation of kanamycin selection on heteroplasmic transformants resulted in vigorously growing, smooth contoured colonies named revertants. This confirmed that the loss of some copies of *ycf* 24 had a dramatic effect on the cell. Resumption of the normal growth rate was shown by the measurement of average colony areas of ragged transformants and WT *Synechocystis* grown with and without kanamycin (Fig. 3.1.4). Average colony areas established that the heteroplasmic state was growth limiting.

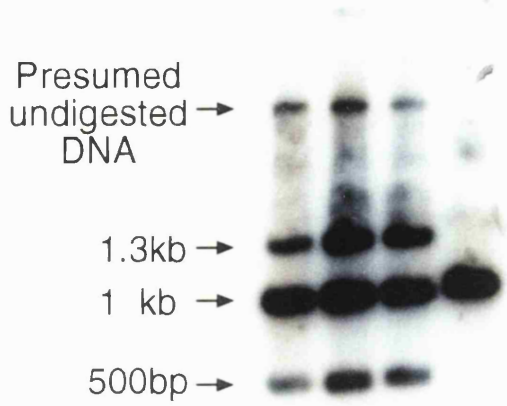
Figure 3.1.2

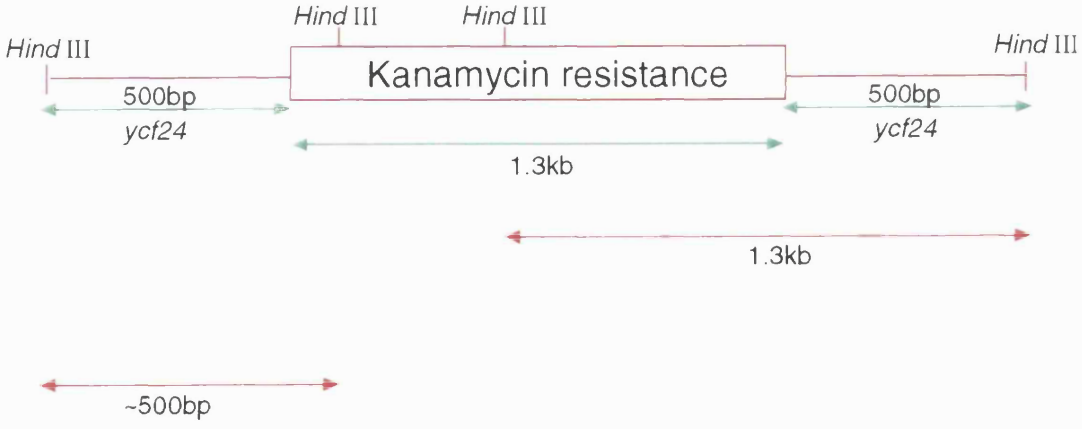
Southern blot of kanamycin resistant transformants using a probe for the central 1kb (*Hind* III) fragment of *ycf* 24.

Lanes 1,2 and 3. Genomic DNA from heteroplasmic mutant clones (two months following the initial selection) showing both the slow growth and ragged phenotypes. The band at 1kb corresponds to WT *ycf* 24 and the bands of 1.3 and 0.5 correspond to the kanamycin disrupted gene.

Lane 4. Wild type *Synechocystis* showing the 1kb band corresponding to WT *ycf* 24 only.

Unlike the WT controls, the lanes corresponding to the heteroplasmic mutant clones, seemed to contain undigested DNA. The difficulty found in preparing DNA from these samples may account for this.


Presumed
undigested
DNA
1.3kb
1 kb
500bp
1 2 3 4


Hind III
Hind III
Hind III
Hind III
Kanamycin resistance
500bp
ycf24
500bp
ycf24
1.3kb
1.3kb
~500bp

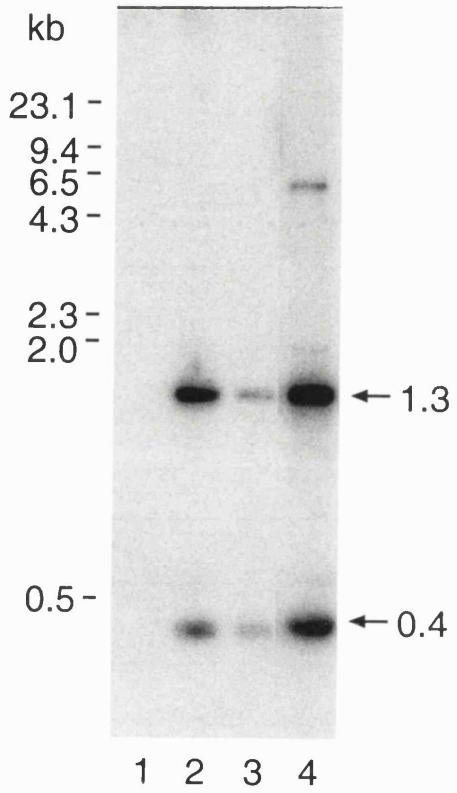


Figure 3.1.3

Southern blot of kanamycin resistant transformants using a probe for the kanamycin resistance gene.

Lane 1. Wild type control showing no hybridization.

Lanes 2, 3 and 4. Heteroplasmic mutant clones containing the 1.3kb band corresponding to the 3′ end of disrupted *ycf* 24 and a fragment of ~400bp corresponding to the internal *Hind* III fragment of the kanamycin resistance gene (see diagram in Figure 3.1.2).

Again some undigested DNA is apparent in lane 4.

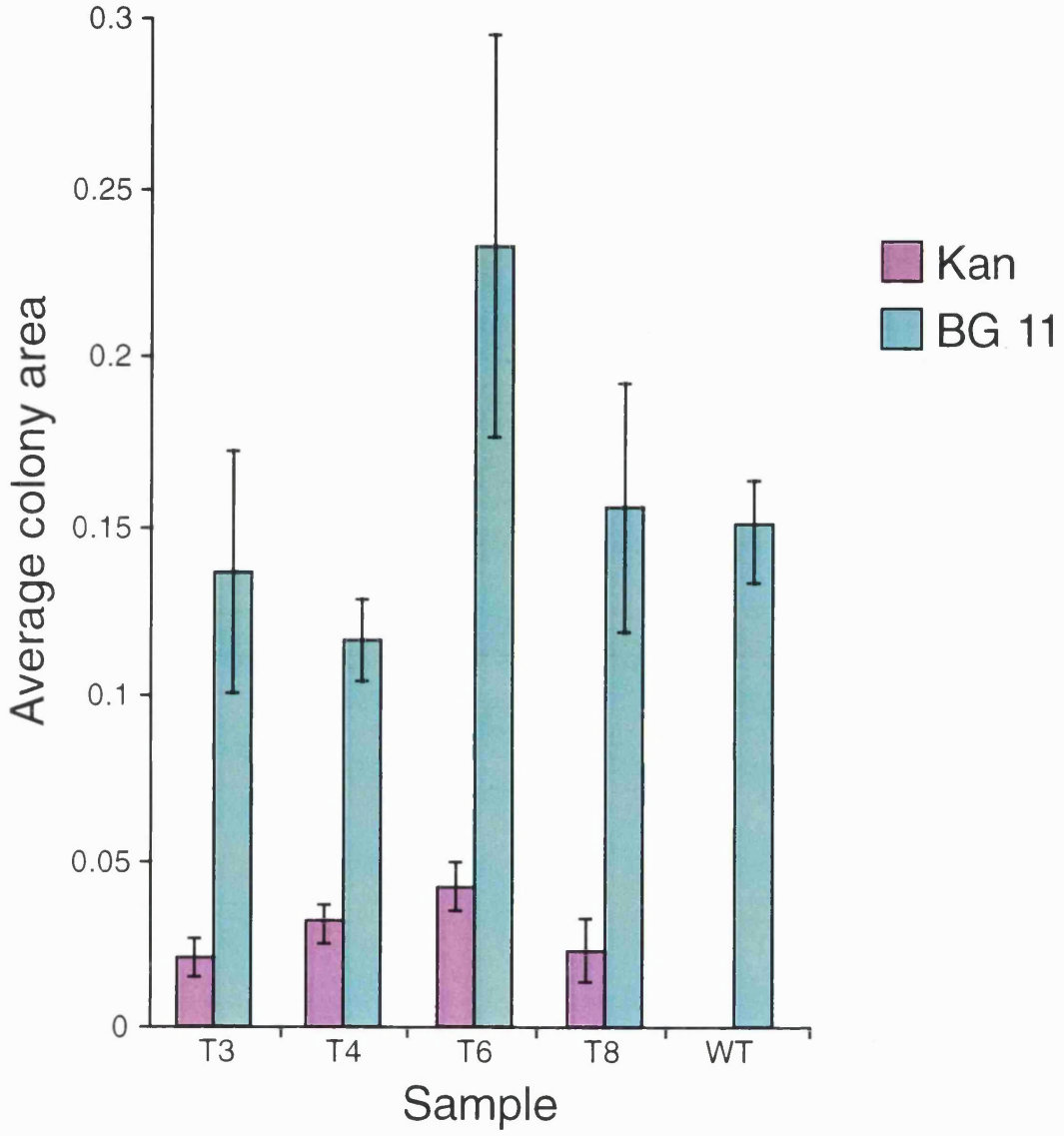


Figure 3.1.4

Average colony areas of four heteroplasmic mutant clones and WT *Synechocystis* grown with and without kanamycin.

The small colony morphology is lost upon the removal of antibiotic selection. Indicating that normal growth rate is resumed. Wild type *Synechocystis* was unable to grow in the presence of kanamycin. The error bars signify the areas of the largest and smallest colonies observed.

Southern blots with DNA from these revertants also were carried out as described previously. These showed that loss of the disrupted gene had occurred as the bands of 1.3 and 0.5kb corresponding to the inactivated gene were absent in one clone and almost absent in another (Fig. 3.1.5). This loss occurred after only 2 weeks growth in the absence of kanamycin.

The first indication of mutant instability was noted at this time. Initially the ragged clones did not grow well in liquid culture, however, fast growing mutants grew out later, a result that will be discussed in detail later (see Section 3.1.4).

After prolonged selection (>12 months), DNA was again extracted from the ragged colonies and more Southern blots were carried out using the WT gene probe for *ycf* 24. Ragged clones remained heteroplasmic, however, the bands corresponding to the disrupted form of *ycf* 24 were less intense than originally shown in Figure 3.1.2 (Fig. 3.1.6). This suggested that selection of fewer, rather than more, copies of the kanamycin disrupted form of *ycf* 24 had occurred over time. The intensity of these bands was analysed using a phosphorimager. Each band was enclosed by a square of uniform size and volume and background volume intensity measurements were calculated automatically. The intensity of each band on the Southern was then determined by subtracting the background volume measurement from the volume measurement. Ratios were calculated on the sum of the intensities for the two bands of the disrupted gene and the wild type band. Ratios for heteroplasmic, ragged mutant clones, mutants grown in the absence of kanamycin, and "smooth" mutants (see from Section 3.1.4) are shown in Table 3.1.1. These data confirmed that the number copies of disrupted *ycf* 24 tended to be lower than the WT gene (Fig. 3.1.6 lanes 2, 3 and 4). It should be noted however, that it was impossible to determine very accurate ratios due to the difficulty experienced in digesting the DNA samples fully. DNA isolated from *Synechocystis* and in particular, from *ycf* 24 disrupted transformants, often had to be digested at 37°C over several days.

Transformants homoplasmic for the disrupted form of *ycf* 24 were never recovered, even after intensive selection with kanamycin at 50 or 100μg/ml. Failure to replace all the endogenous copies with the disrupted form of *ycf* 24 even after prolonged selection is a strong indication that gene disruption was lethal (Golden, 1988), suggesting that the reduced colony growth rate and ragged phenotype may be attributed to the death of homoplasmic cells.

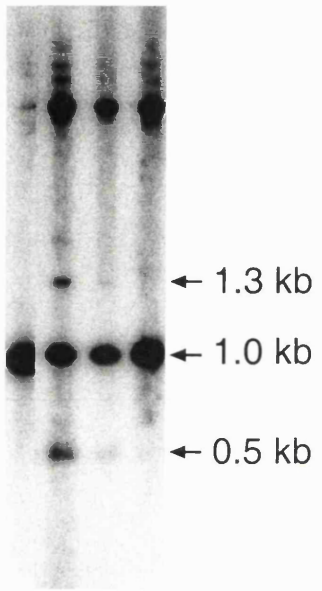


Figure 3.1.5

Southern blot using a WT gene probe for *ycf* 24 of heteroplasmic mutants carrying disrupted *ycf* 24 grown in the absence of kanamycin.

Lane 1. Wild type control showing the 1kb fragment corresponding to the WT gene only.

Lane 2. Heteroplasmic ragged mutant showing the 1.3 and 0.5 bands corresponding to disrupted *ycf* 24 as well as the WT band (positive control).

Lanes 3 and 4. Heteroplasmic mutant clones showing the loss of the 1.3 and 0.5kb bands corresponding to disrupted *ycf* 24. This is more apparent in lane 3 than lane 2 where complete loss has occurred indicating the disrupted gene can be selected out of the cell population. The band remaining corresponds to the 1kb WT gene fragment.

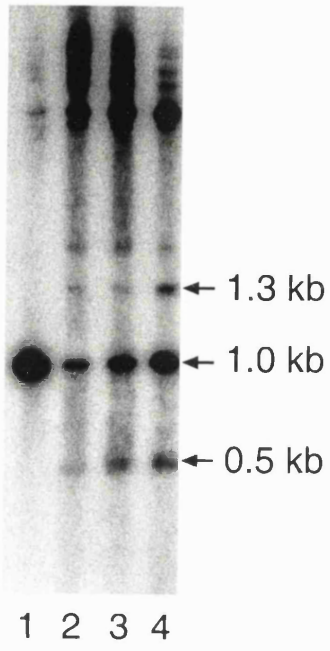


Figure 3.1.6

Southern blot of heteroplasmic mutant clones selected for over a year on kanamycin using a WT gene probe for *ycf* 24.

Lane 1. Wild type control.

Lanes 2, 3 and 4. Heteroplasmic ragged mutant clones showing that the 1kb band corresponding to the WT gene is more intense than the sum of the 1.3 and 0.5kb bands corresponding to disrupted *ycf* 24. This was unlike the results seen in the original Southern blots, an example of which is Figure 3.1.2. However, a considerable Amount of undigested DNA is present in these preps.

	Volume - background volume		
Cell type	Wild type *ycf*24	Disrupted *ycf*24	Ratio WT: disrupted *ycf*24
WT	853588	N/A	N/A
T3	71462	80690	1:1.1
T8	179522	87050	1:0.5
O8	270705	106542	1:0.4
T3 - Km	192531	29557	1:0.1
SM3 50	508030	597254	1:1.2
SM3 100	517561	563135	1:1.1
SM8 50	2066984	2233694	1:1.1
SM8 100	781137	543489	1:0.7

Table 3.1.1

Ratio of wild type to kanamycin disrupted forms of *ycf* 24 in *Synechocystis* transformants calculated using phosphorimaging data.

WT: Wild type *Synechocystis*.

T3, T8 and O8: Heteroplasmic ragged mutant clones.

T3 - Km: Hereroplasmic ragged mutant grown in the absence of kanamycin (revertants).

SM3 50, SM3 100, SM8 50 and SM8 100: Heteroplasmic smooth mutants grown in the presence of 50 of 100μg/ml kanamycin.

3.1.2 Disruption of *ycf*24 in *Synechocystis* with the *aadA* resistance cassette

To eliminate the possibility of some non-specific effect of kanamycin and to investigate whether disruption with a different resistance marker would give the same result, the kanamycin resistance gene was replaced with the *E. coli* gene *aadA* (Sp1) which confers resistance to both streptomycin and spectinomycin (Hollingshead and Vapnek, 1985; Goldschmidt-Clermont, 1991). The *E. coli* gene, cloned into the *Hind* III site of bluescript (PBSHDSp1), was provided by Mark Ashby (UCL). After digestion with *Hind* III, klenow treatment and purification, Sp1 was cloned into *ycf*24 and used to transform WT *Synechocystis*. Again inactivation mutants (although not checked by Southern blots) had the same mutant colony morphology. These experiments confirmed the original result, showing that the mutant phenotype was reproducible with a different selective agent.

3.1.3 Growth of *Synechocystis ycf*24 disrupted mutants on supplemented media

Having confirmed that disruption of *ycf* 24 produced a mutant phenotype, it was considered whether the phenotype could be linked to the inability of the mutants to carry out some biosynthetic pathway. The possibility of reversion to wild type phenotype was investigated by supplementing the BG11 growth medium with aromatic or aspartate amino acids. The reasoning behind this was that plastids in plants play an important role in the biosynthesis of these two groups of essential amino acids (Lea *et al.*, 1982) and some such metabolic role could explain the persistence of *ycf*24 in apicomplexans.

Plates containing BG11 supplemented with 0.1mM of each of three aromatic amino acids (L-phenyalanine, L-tryptophan and L-tyrosine) or two aspartate amino acids (L-threonine and L-lysine) were streaked with ragged mutants and wild type *Synechocystis*. A similar experiment was set-up using nitrite as opposed to nitrate as a source of nitrogen in the BG11 medium. A source of nitrogen is paramount to essential amino acid biosynthesis and most cyanobacteria use either nitrite or ammonium ions as a nitrogen source. An exception are the nitrogen-fixing cyanobactera (Flores and Herrero, 1994) and although *Synechocysis* is not one of these, a gene homologous to a nitrogen fixation gene, *nifS*, is found downstream of *ycf* 24 and *ycf* 16. None of these tests indicated any improvement in growth rate or change in colony morphology.

3.1.4 Instability of ycf24 heteroplasmic mutants in liquid culture and reversion to WT phenotype

After a few months of selection an attempt was made to grow the ragged transformants in liquid culture so that larger quantities of DNA could be extracted for Southern blots. Initially the growth rate was poor as is common with *Synechocystis* mutants (C. Mullineaux, personal communication). After about six weeks however, (with kanamycin being added every week or so) the mutants in liquid culture began to grow as well as WT liquid cultures. When samples from these cultures were streaked on to BG11 agar supplemented with 50μg/ml of kanamycin the transformants no longer showed the ragged colony morphology and growth pattern. Instead growth was as prolific as the wild type and colonies were smooth. Initially it was thought that this phenotype was due to the instability of kanamycin under these growth conditions or a defective batch of the antibiotic, i.e. if selective pressure had been lost the inactivated form of *ycf* 24 could also have been lost from the cell population, resulting in selection of the WT phenotype. To test this, fresh liquid cultures were established from these colonies (for mutant clones 3 and 8) with double the concentration of freshly prepared kanamycin (100μg/ml) and allowed to grow for a few weeks, subculturing every week with fresh kanamycin to maintain selective pressure. When samples of these liquid cultures were streaked onto solid selective medium the vast majority of colonies were still round and smooth. A test plate onto which WT *Synechocystis* had been streaked confirmed the antibiotic was effective as no growth was apparent. Southern analysis using a probe for wild type *ycf* 24 was carried out on DNA isolated from three samples of these fast growing, smooth contoured, kanamycin resistant mutants. This revealed that all of the clones remained heteroplasmic, containing copies of both WT *ycf* 24 and kanamycin inactivated *ycf* 24, with a ratio of WT:disrupted *ycf* 24 of approximately 1:1 (Fig. 3.1.7). This reversion of mutant clones grown (for some time) in liquid culture to wild type growth led us to call them smooth suppressants. One possible explanation for "suppression" reversed the hypothesis that the copy number of wild type to inactivated form of *ycf* 24 determines the phenotype. In other words, when single cells segregate, the ratio of inactivated *ycf* 24 to wild type *ycf* 24 in each cell is likely to be different because *Synechocystis* contains multiple copies of *ycf* 24. It is possible therefore that cells with

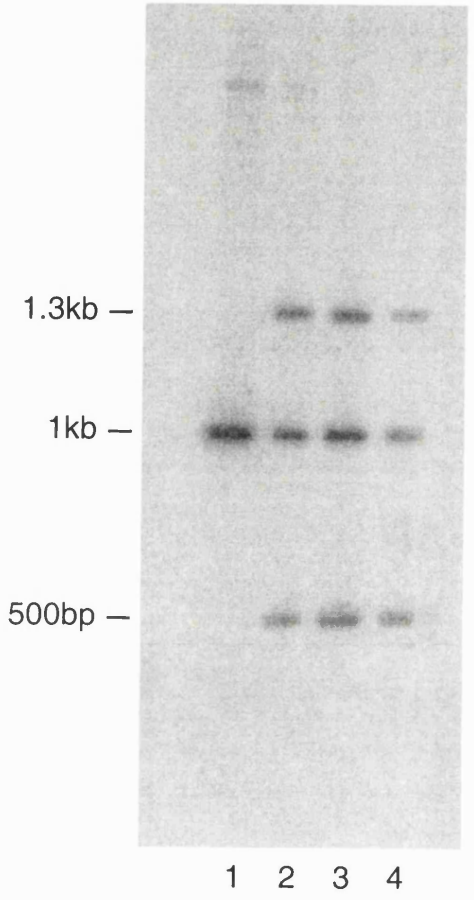


Figure 3.1.7

Southern blot using a WT gene probe for *ycf* 24 on DNA isolated from smooth kanamycin resistant mutants of *Synechocystis*.

Lane 1. WT control.

Lanes 2, 3 and 4. Smooth mutants showing the 1.3 and 0.5kb bands corresponding to kanamycin disrupted *ycf* 24 as well as the 1kb band corresponding to the WT gene, indicating they had remained heteroplasmic.

a low ratio of mutant to wild type genes could enable them to grow as wild type, whereas a higher ratio was either lethal or caused abnormalities in growth. As Figure 3.1.6 and Table 3.1.1 (mentioned in Section 3.1.1) show, there tended to be fewer (approximately half) copies of the disrupted gene compared to the wild type gene in heteroplasmic clones with the mutant or ragged phenotype. By contrast, in the smooth mutants, the ratio of wild type to inactivated gene increased to approximately 1:1. These data are shown in Table 3.1.1 for two smooth mutants grown in the presence of either 50 or 100μg/ml of kanamycin. As the ratio of WT to disrupted *ycf* 24 was higher in these smooth mutants than in the ragged heteroplasmic mutants (indicating that these mutants were able to support a greater reduction in WT *ycf* 24 and still seem healthy) we suspected a compensatory suppressor mutation had occurred elsewhere in the genome.

Support for a complementing mutation was found when one of three liquid cultures derived from mutant clone T3 (a ragged transformant from a second round of transformations) apparently became homoplasmic for the inactivated *ycf* 24 gene under kanamycin selection (Fig. 3.1.8, lanes 1 and 2). Not only was this isolate viable when streaked on BG11 plates supplemented with kanamycin, but gave large, round and smooth colonies like wild type. By contrast another isolate, shown to be heteroplasmic for the inactivated form of *ycf* 24, gave much smaller, ragged colonies (lanes 3 and 4). If, according to the earlier hypothesis, deletion of *ycf* 24 was lethal, homoplasmic mutants could survive only if a compensating mutation had occurred. Subsequent Southern blots forced a re-interpretation of this liquid culture result. The clone originally thought to be homoplasmic was found in fact to be heteroplasmic (Fig. 3.1.9 lanes 2 and 3).

As the original liquid cultures were grown under very high concentrations of kanamycin (100μg/ml), as soon as the kanamycin concentration was reduced (by streaking on plates to obtain colonies) it is possible that the kanamycin-disrupted copies of *ycf* 24 were subject to less heavy selection, allowing the WT form again to be observed on Southern blots. Thus, I may have originally had a heteroplasmic clone with very few copies of WT *ycf* 24 which were undetectable by Southern analysis.

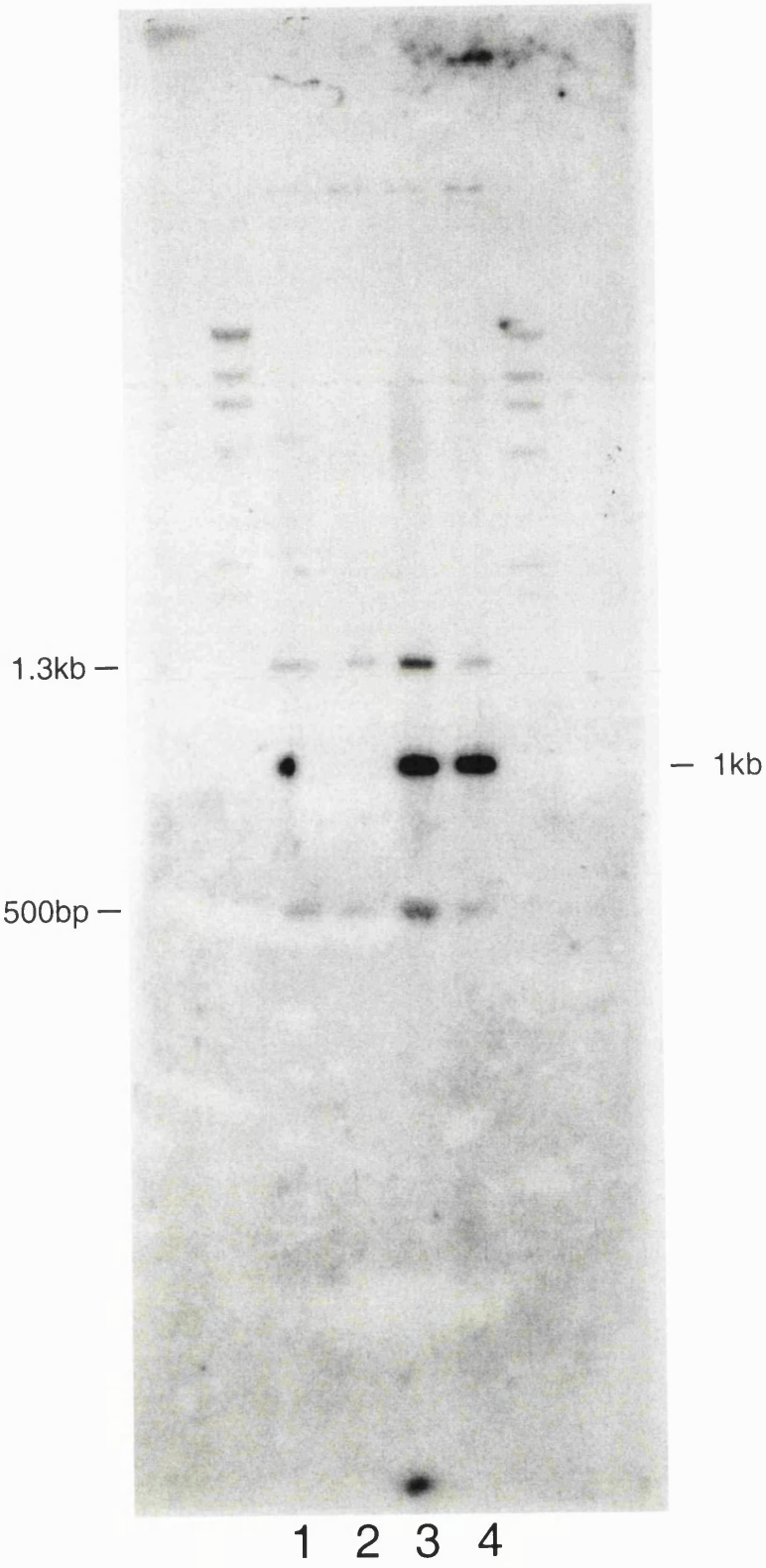
Alternatively, the transfer of DNA from the agarose gel onto the nitro-cellulose membrane around the position of the WT gene may not have occurred correctly giving an artefactual appearance of homoplasmy.

Figure 3.1.8

Southern blot using a WT gene probe for *ycf* 24 on DNA isolated from two isolates of the ragged mutant T3 grown in liquid culture.

Lanes 1 and 2. Only faint bands for the disrupted form of *ycf* 24 (1.3 and 0.5kb) were found. When grown on solid media this isolate gave rise to rapidly growing, smooth contoured colonies. However, a signal of unusual shape is apparent in lane 1 around the position of the WT gene. This may signify an air bubble leading to the mis-interpretation of homoplasmy.

Lanes 3 and 4. Ragged heretoplasmic mutant clone containing both the disrupted form of *ycf* 24 and the WT gene (1kb band).


1.3kb
1kb
500bp
1 2 3 4

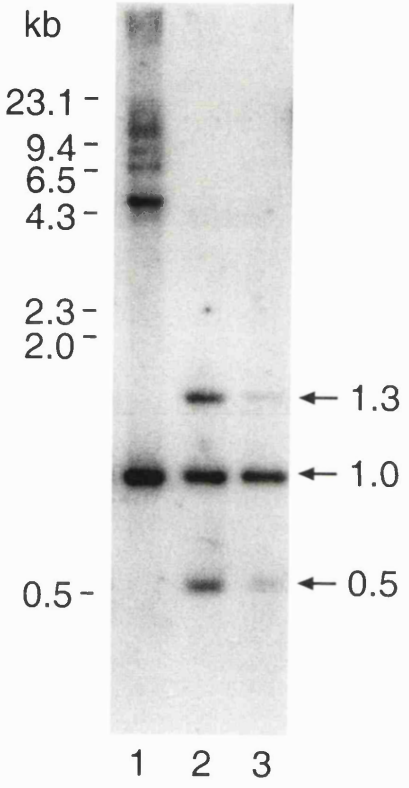


Figure 3.1.9

Southern blot of an isolate of *Synechosystis* mutant T3 previously thought to be homoplasmic for the disrupted form of *ycf* 24.

Lane 1. Wild type control.

Lanes 2 and 3. Isolate of T3. Bands of 1.3, 1 and 0.5kb show that the isolate is heteroplasmic, containing both copies of the inactivated and WT forms of *ycf* 24.

3.1.5 Transforming WT *Synechocystis* with DNA from smooth mutant clone 8

To further investigate the question of complementation it was decided to transform wild type *Synechocystis* with DNA isolated from the smooth mutants. Also an attempt was made to complement the ragged mutant with DNA from smooth mutants.

Transformation of wild type cells was carried out with approximately 1μg of total genomic DNA from one of the smooth mutant clones (clone 8, previously shown to be heteroplasmic by Southern analysis). Following overlay with 50μg/ml of kanamycin, hundreds of transformants were observed. This time the transformants did not appear to be ragged, and colonies were as large as the wild type. This provided evidence for some kind of complementing mutation. However, it was expected that a mixture of both smooth and ragged colonies would be seen unless all the chromosomes carrying the disrupted gene also carried the compensatory mutation.

Hybridization to *Hind* III restriction fragments of DNA isolated from three of these transformants using labelled *ycf* 24 (see Fig. 3.1.10), showed they lacked the 3' end of the gene, i.e. the 1.3Kb fragment of *ycf* 24 (Fig. 3.1.10A, lanes 1, 2 and 3) previously seen in all kanamycin inactivated *ycf* 24 mutants. (Lane 4 in Fig. 3.1.10A corresponds to smooth clone 8 and includes the 1.3Kb *Hind* III restriction fragment and lane 5 is the wild type control). Hybridization of a kanamycin gene probe to the same restriction fragments indicated that the kanamycin gene had recombined elsewhere into the genome (Fig. 3.1.10B). Again the 1.3Kb band was absent from lanes 1, 2 and 3 but not lane 4 (no hybridization to wild type DNA, lane 5, occurred). In addition, another restriction fragment was highlighted between the 6.5 and 4.3Kb markers that was not evident with the probe for WT *ycf* 24. These data indicated that recombination of the resistance gene with loss of the 3′ sequence for *ycf* 24 had occurred. Because recombination resulted in the insertion of approximately only half of *ycf* 24 it is possible that there had been a mis-match with part of the disrupted *ycf* 24 recombining into the genome in the wrong place without replacing the wild type gene. Thus cells would be resistant to kanamycin and all copies of the genome would contain the WT gene. It follows that without inactivation there would be no ragged phenotype.

When this experiment was repeated for a second and third time the results were more as expected. Although some colonies were smooth contoured and large,

Figure 3.1.10

A. Southern blot of *Synechocystis* transformants using a WT gene probe for *ycf* 24 following transformation of WT cells with DNA isolated from smooth mutant clone 8.

Lanes 1, 2 and 3. Kanamycin resistant transformants. The bands of 1kb and 0.5kb suggest that the 3′ end of *ycf* 24 had not recombined into the genome along with the kanamycin resistance gene but has been lost.

Lane 4. Smooth mutant (clone 8) control which contains the 1.3kb band.

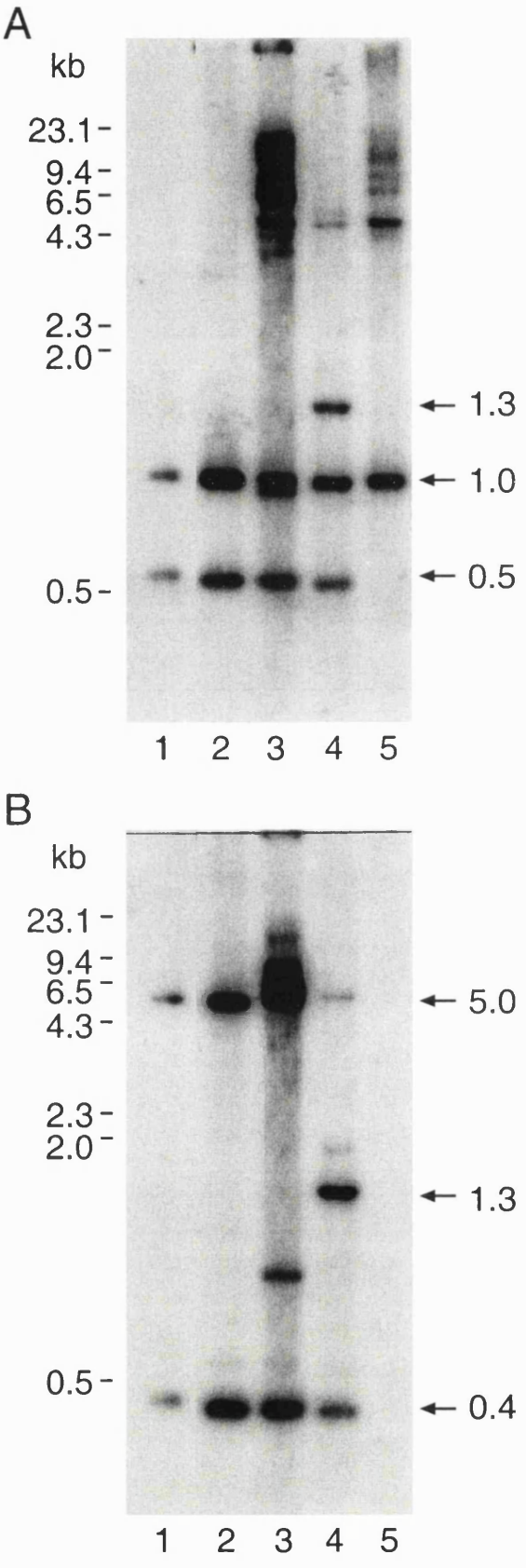
Lane 5. Wild type control shows the expected band at around 1kb.

B. Southern blot of *Synechocystis* transformants using a kanamycin gene probe for *ycf* 24 following transformation of WT cells with DNA isolated from smooth mutant clone 8.

Lanes 1, 2 and 3. Kanamycin resistant transformants. The missing 1.3kb band and additional band of approximately 5kb confirmed the absence of the 3′ end of *ycf* 24 with the retention of the kanamycin resistance gene.

Lane 4. Smooth mutant (clone 8) control which contains the 1.3kb band.

Lane 5. Wild type control showing no hybridization had occurred.


A
kb
23.1
9.4
6.5
4.3
2.3
2.0
0.5
1.3
1.0
0.5
1 2 3 4 5
B
kb
23.1
9.4
6.5
4.3
2.3
2.0
0.5
5.0
1.3
0.4
1 2 3 4 5

like wild type, ragged colonies were also apparent. This indicated segregation of the compensatory mutation. When these smooth and ragged colonies were further selected by streaking onto fresh BG11 plates supplemented with 50μg/ml of kanamycin, smooth colonies tended to give rise to smooth colonies (the odd ragged one was observed) whereas ragged colonies gave rise to more varied colony sizes and shapes. Some ragged clones resulted in only small ragged progeny, others in larger ragged colonies. The occasional large, smooth colony was also observed. These results suggested that selection of the mutation was occurring in some of the clones causing reversion of their phenotype from ragged to smooth.

3.1.6 Complementation of the ragged phenotype with the smooth phenotype

DNA from three smooth mutants (clones 3, 4 and 8) was used to transform three ragged transformants T3, T4 and T8 by direct plating onto BG11 plates containing kanamycin. After approximately ten days growth, colonies viewed under the microscope showed a small percentage (between 1 and 2 %) of colonies on each of the plates were larger and smoother than others which were small and ragged. Even so, these colonies were not as large or smooth as the wild type cultures suggesting that further selection of the compensatory gene was required.

When a sample of smooth colonies were picked and re-streaked onto fresh BG11 plates containing kanamycin the results were as observed in the previous experiment; i.e. with one or two exceptions, the resulting colonies were smooth. The ragged colonies however, tended to give rise to ragged progeny only, unlike the wild type transformation. These results indicate that only a few copies of the compensating gene had been selected in the original transformation.

It should be mentioned that the control transformations which included the transformation of DNA from WT into WT *Synechocystis* cells, DNA from smooth mutants into smooth mutants and DNA from ragged mutants into ragged mutants gave the expected results, i.e. WT and smooth mutant colonies remained large and smooth, whereas ragged mutants remained ragged.

3.1.7 Mutant rescue

The partial knockout of *ycf* 24 in *Synechocystis* with a kanamycin disrupted gene resulted in a number of mutant phenotypes including slow growth and a ragged

colony morphology (others will be discussed in the next chapter). Although it had been shown that some or all of these phenotypes could be lost by either compensatory mutations (smooth mutants) or by the removal of kanamycin pressure ("revertants"), it was decided to carry out a gene rescue experiment. The strategy was to replace the disrupted gene with a wild type gene to which an antibiotic resistance gene had been tagged. If this replacement resulted in the rescue of the WT phenotype it would confirm that the phenotypes were directly caused by the partial knockout of *ycf* 24.

To make this construct (see Fig. 3.1.11), the bluescript vector pBSHdSp1, containing the streptomycin resistance gene (see Section 3.1.2) was initially digested with *Bam* HI. This enzyme site occurs 33bp downstream of the *Hind* III site within which the *aadA* resistance gene had been cloned. Following amplification of full length *ycf* 24 from *Synechocystis* PCC6803 (using primers designed with *Bam* HI linkers) *ycf* 24 was cloned into this site, i.e. 33bp downstream of Sp1. Recombinant clones were selected in INVαF′ cells by growth on streptomycin and analysed for both the *ycf* 24 and Sp1 inserts by digestion with *Bam* HI and *Hind* III. Failure of *ycf* 24 to clone into the *Bam* HI site of pBSHdSp1 would result in a single fragment of ~4.5kb, corresponding to the size of pBSHdSp1 upon restriction with *Bam* HI. Digestion with *Hind* III on the other hand would result in two bands, one of ~3kb, corresponding to pBSKS+ and another of ~1.5kb corresponding to the streptomycin resistance gene if cloning had failed. Successful cloning of *ycf* 24 resulted in an additional band of 1.5kb (*ycf* 24) upon digestion with *Bam* HI and a total of four bands (of different sizes depending upon the orientation) after digestion with *Hind* III.

Thirty clones resistant to streptomycin were analysed. Only four were found to have *ycf* 24 cloned into the *Bam* HI site. Two of these clones were sequenced and used to transform both WT *Synechocystis* and ragged *ycf* 24 mutants which were overlaid with streptomycin after four days. Although there was no flanking sequence homologous to *Synechocystis* DNA on either side of the streptomycin resistance gene, it was hoped that it would recombine into the *Synechocystis* genome following the recombination of WT *ycf* 24 because of its selection and close proximity. Unfortunately, after two weeks growth on streptomycin, no streptomycin resistant transformants were observed from either transformation of WT or ragged *ycf* 24 mutant cells.

Figure 3.1.11

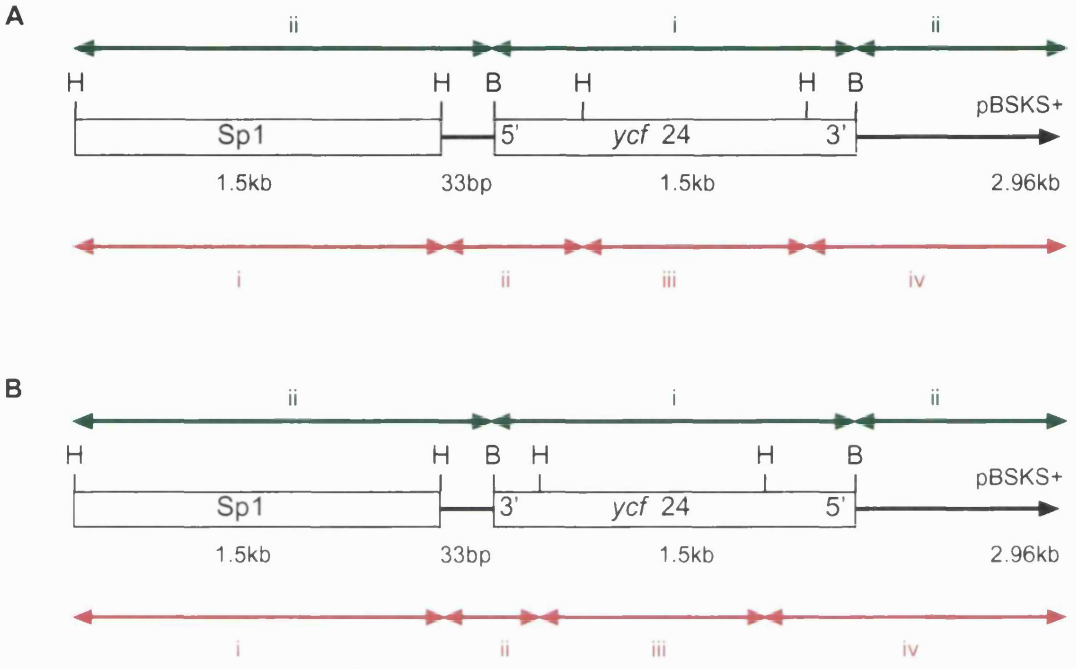
Sizes of restriction fragments expected following cloning of *Synechocystis ycf* 24 into the *Bam* HI site of pBSKS+.

Orientation 1

Hind III sizes expected (approximate)		*Bam* HI sizes expected (approximate)
I.	1.5kb (Sp1)	1.5kb (*ycf*24)
II.	330bp	4.5kb (pBSHdSp1)
III.	1.0kb	
IV.	3.14kb	

Orientation 2

Hind III sizes expected (approximate)		*Bam* HI sizes expected (approximate)
I.	1.5kb (Sp1)	1.5kb (*ycf*24)
II.	210bp	4.5kb (pBSHdSp1)
III.	1.0kb	
IV.	3.3kb	


A
ii
i
ii
H
H
B
H
H
B
pBSKS+
Sp1
5'
ycf 24
3'
1.5kb
33bp
1.5kb
2.96kb
i
ii
iii
iv
B
ii
i
ii
H
H
B
H
H
B
pBSKS+
Sp1
3'
ycf 24
5'
1.5kb
33bp
1.5kb
2.96kb
i
ii
iii
iv

The experiment was repeated twice but always failed. This was likely to be due to recombination of the WT *ycf* 24 gene only. Ideally the experiment should have been repeated with the streptomycin resistance gene inserted into a cloned *Synechocystis* sequence, either up-stream or down-stream of *ycf* 24, leaving enough homologous sequence on either side to allow recombination to occur. However, it is not known whether disruption of such flanking sequences would have caused any adverse affects.

3.1.8 Conclusion

The partial knockout of *Synechocystis ycf* 24 maintained during long term selection suggests that homoplasmy is lethal and that *ycf* 24 is an essential gene. Hetreroplasmic cells giving rise to slow growing, ragged contoured colonies did not revert to WT on supplementing the growth medium with a small selection of essential amino acids. However, the slow growth and ragged phenotypes was reversed by compensatory mutation(s) that occurred spontaneously elsewhere in the genome. As none of these observations gave any indication of the role of *ycf* 24 in *Synechocystis* PCC6803, further investigation of the characteristics of the ragged and smooth *Synechocystis* transformants was undertaken, as discussed in Chapter 3.2.

3.2 Morphology and characteristics of *Synechocystis* transformants

3.2.1 Transmission electron microscopy of cell sections

Transmission electron microscopy (carried out by Elizabeth Hirst at NIMR) highlighted differences in the fine structure of the transformed cells. In the WT (Fig. 3.2.1), thylakoid membranes (shown by arrow) were separated by electron dense phycobilisomes, large protein complexes consisting of phycobilisome proteins carrying the major light harvesting pigments of cyanobacteria. By contrast, between the thylakoid membranes of the ragged mutant cells (Fig. 3.2.1B and C) there appeared to be small electron opaque dots surrounded by an electron transparent area ("plaque" shown by double arrow). Mutant cells plated in the absence of kanamycin reverted back to WT morphology, with only 45% of cells showing plaques as opposed to 79% when grown with kanamycin (Table 3.2.1). This raised the possibility that *ycf* 24 might code for a phycobilisome associated protein. As phycobilisome proteins are broken down under conditions of amino acid shortage or nitrogen limiting conditions (Stanier, 1988) this might indicate again that *ycf* 24 could have a role in amino acid biosynthesis or nitrogen assimilation. However, as explained in the previous chapter, experiments to complement growth with various amino acids and a different nitrogen source were unsuccessful (Section 3.1.3). Alternatively, the plaques result from stored glycogen or even liquid droplets (Stanier, 1988). The former are characteristically found in cells lacking nitrogen. The possibility also was raised (L. Hirst, NIMR, personal communication), that the electron dense areas were viruses which may be dormant in wild type cells. According to this scenario, *slr*0074 (*ycf* 24) might code for a protein that protects the cell from viruses. Distinct from these speculations, perhaps the most relevant observation however, is that 4% of cells studied from a WT culture also had some plaques like the mutant phenotype (Table 3.2.1). Thus the plaques are likely to be diagnostic of sick cells rather than a primary effect of the gene disruption.

Smooth mutants also contained electron opaque plaques. In addition a small number of these cells were also mis-shapen, appearing contorted with multiple indentations on the surface suggesting the cells had begun to divide but failed to complete cytokinesis. Examples of a few of these cells are shown in Figure 3.2.2.

Figure 3.2.1

Transmission electron micrographs of cell sections of wild type *Synechocystis* and *ycf*24 mutants.

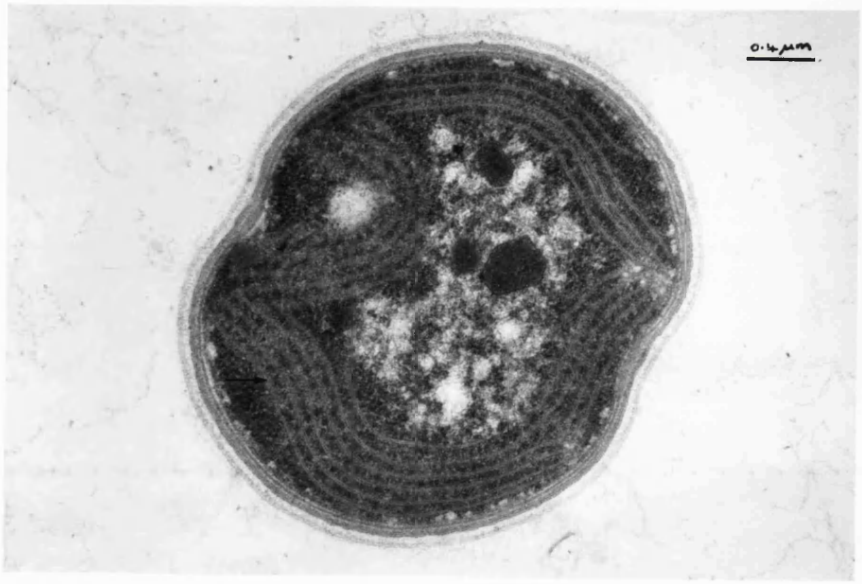
A. Wild type *Synechocystis* PCC6803. The single arrow highlights the regular, stacked appearance of the thylakoid membranes.

B. Heteroplasmic ragged mutant clone T3.

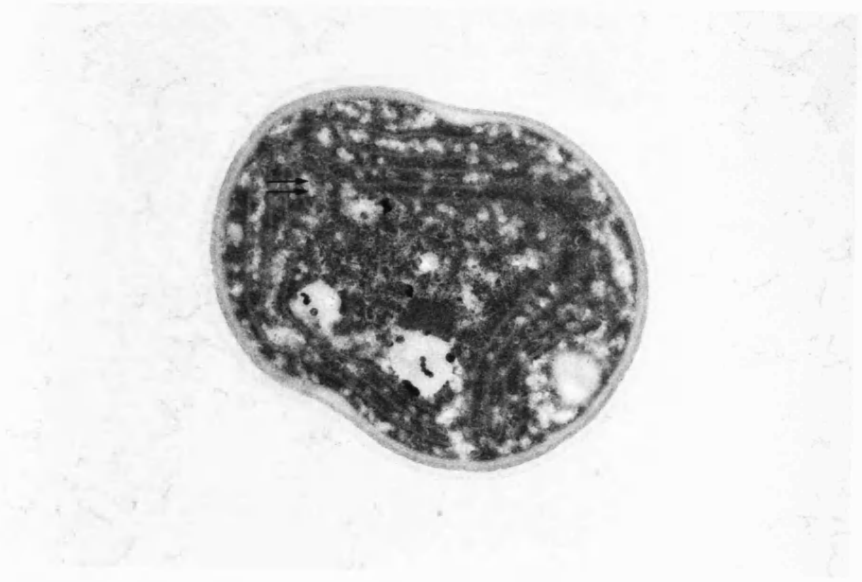
C. Heteroplasmic ragged mutant clone T8.

The double arrows in both B and C highlight the electron opaque dots surrounded by a transparent area, plaques, occurring between the thylakoid membranes.

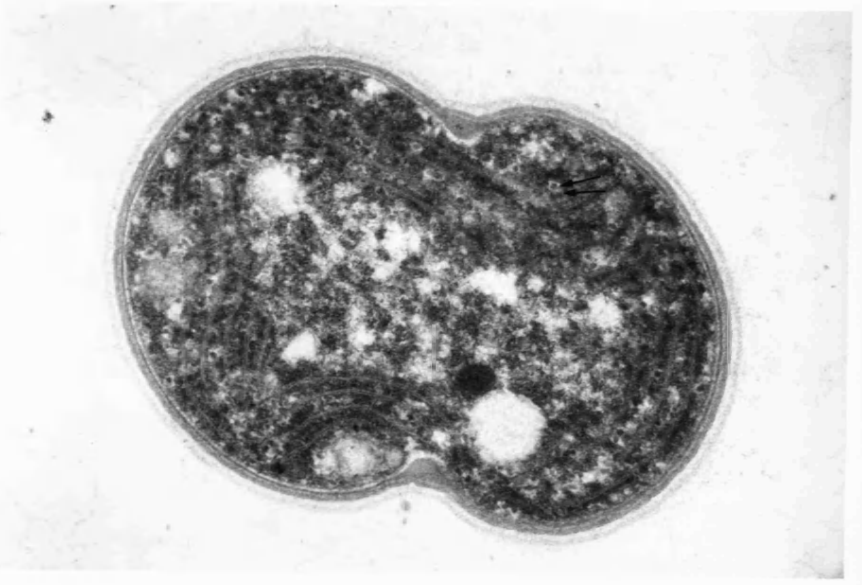
A



B



C



Cell type	Total no. of cells	% Wild type	% Mutant (plaques)
WT	112	96	4
T3 and T8	175	21	79
T3 and T8 - Km	176	55	45

Table 3.2.1

Total number and percentage of cells showing either the WT or mutant phenotype from a sample of WT *Synechocystis* and ragged *ycf* 24 mutants grown with or without kanamycin.

WT: Wild type Synechocystis.

T3 and T8: Heteroplasmic ragged mutant clones.

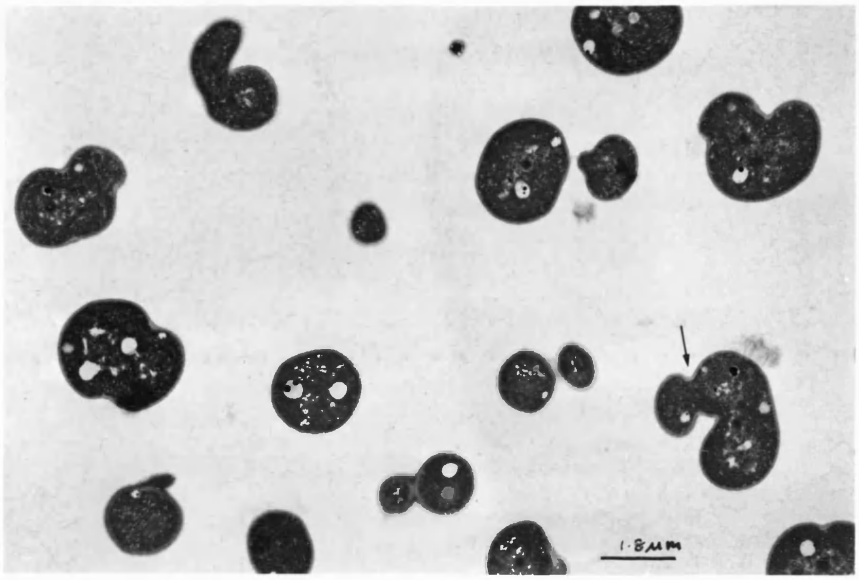
T3 and T8 - Km: Heteroplasmic ragged mutants grown in the absence of kanamycin (revertants).

Figure 3.2.2

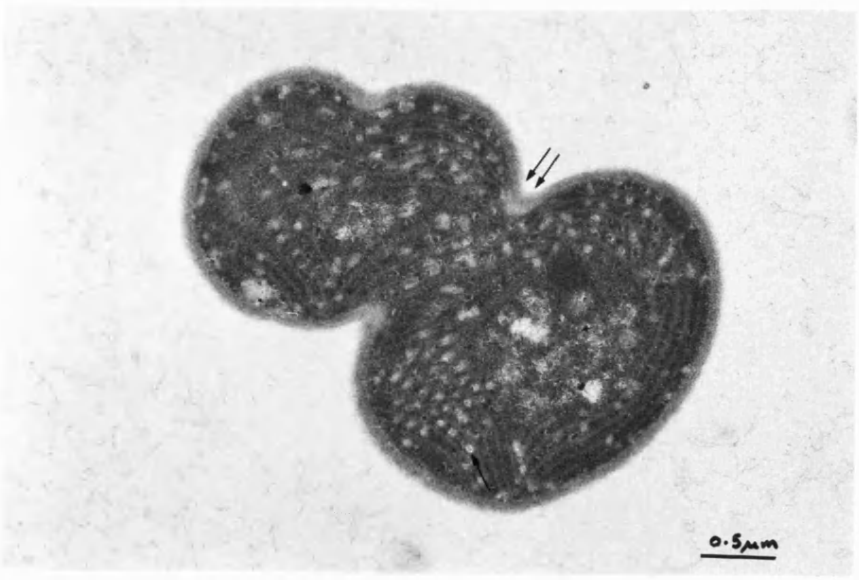
Transmission electron micrographs of cell sections from a sample of smooth mutant clone 8 (SM8).

Although the plaques characteristic of the ragged mutants are present (single arrow), mis-segregation is also apparent. Some cells appeared distorted, as in both A and B, with indentations on the surface (double arrows) indicating a failure to complete cytokinesis.

A



B



3.2.2 Scanning electron microscopy of cell surfaces

Scanning electron microscopy was undertaken both by Liz Hirst (NIMR) and myself with a sample of wild type *Synechocystis* and two samples of ragged mutant clone T8, one grown under continuous kanamycin pressure (50μg/ml) and the other allowed to grow in the absence of kanamycin for two weeks. These observations showed a clear phenotypic difference and indicated a possible role for *ycf*24 in DNA segregation and/or cell division. It was found that mutant cells grown in the presence of kanamycin were much more likely (55%) to be in the process of cell division (Table 3.2.2). Only 12-19% of cells were found to be undergoing cell division in either the wild type (Fig. 3.2.3A) or "revertant" samples. Also, many of the dividing cells from the ragged mutant clone grown on kanamycin appeared to be in a late stage of septation (Fig. 3.2.3B). Two unseparated daughter cells were visible in approximately 34% of these cells, unlike either the wild type or ragged mutants removed from kanamycin pressure, where only 8% and 4% respectively of cells were in the late stages of division. Chi-square analysis shows these differences were significant, i.e. P (probability) values of less than 0.001 (Table 3.2.2). Most of smooth mutants also were found to be at a late stage of cytokinesis, with occasional small chains of incompletely separated cells (Fig. 3.2.3C and Table 3.2.2). These data coupled with the transmission electron microscopy results indicate that suppression of the mutation by complementation may have increased the frequency of septation.

To confirm that the high numbers of cells seemingly undergoing cytokinesis had not yet separated, a sample of ragged mutants was subjected to a short pulse of sonication before observing under the microscope. The proportion of dividing cells remained the same confirming the partial knockout of *ycf*24 in *Synechocystis* affected cell division (cytokinesis/septation).

3.2.3 DNA content of *Synechocystis* transformants

Samples of *Synechocystis* cells were washed in 1 x PBS and attached to microscope slides by 1x poly-L-lysine. Following fixation by methanol and air drying DNA was stained by incubating with DAPI for approximately 30 seconds. After rinsing with distilled water and air drying the stained samples were covered with Vectashield to prevent quenching of nucleoid staining when viewed under UV light microscopy. This work was done by Iain Wilson (NIMR).

	% Fission			
Cell type [*1]	Early stage	Late stage	Total	$P(\chi^2)$ [*2]
WT	10	9±1	19	-
T3 and T8	21±6	34±1	55	<0.001
T3 and T8 - Km	8	4	12	0.9
SM8	20	41	61	<0.001

Table 3.2.2

Comparison of the divisional state shown by scanning electron microscopy of WT *Synechocystis* and *ycf*24 mutants.

WT: Wild type *Synechocystis.*

T3 and T8: Heteroplasmic ragged mutant clones.

T3 and T8 - Km: Hereroplasmic ragged mutants grown in the absence of kanamycin (revertants).

SM8: Heteroplasmic smooth mutant clone.

[*1]Number of cells counted >100/group (mean from 2 experiments).

[*2]Comparison with WT.

Figure 3.2.3

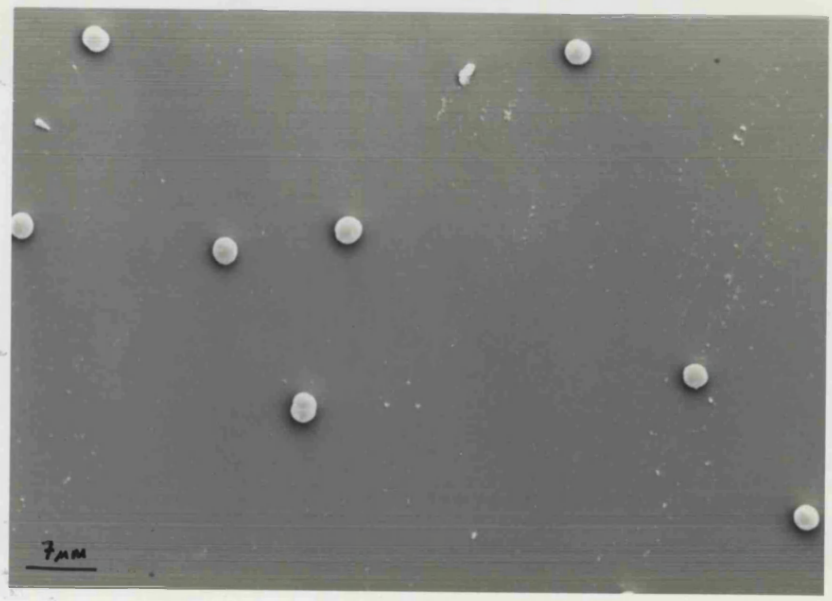
Scanning electron micrographs of wild type *Synechocystis* and *ycf*24 mutants.

A. Wild type. Most of the cells are not undergoing cell division.

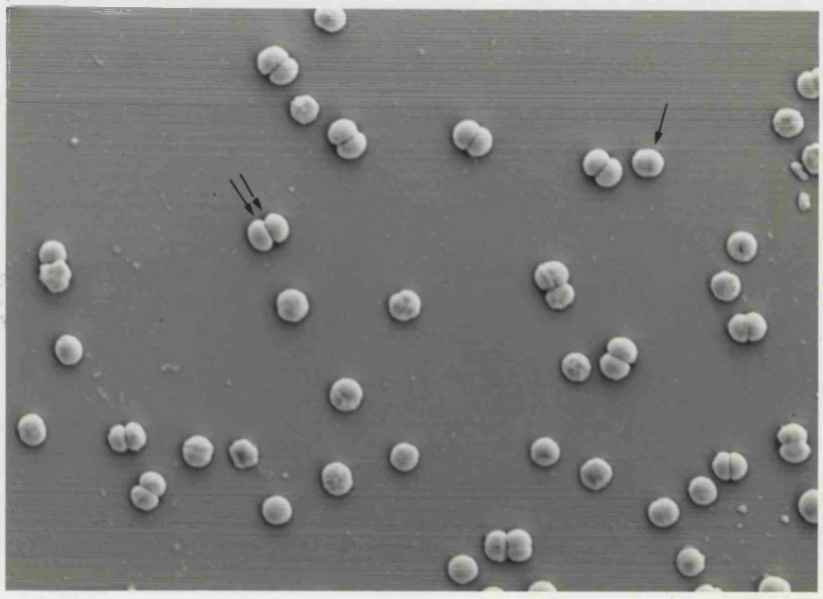
B. Heteroplasmic ragged mutant clone T8. Many of the cells are undergoing cell division. Some are in the early stages (as shown by the single arrow) however the majority appear to be in the late stages where two distinct cells can be seen. Examples of these are shown by the double arrows.

C. Heteroplasmic smooth mutant clone SM8. Again many of the cells appear to be undergoing cell division, particularly the late stages as shown by the single arrows.

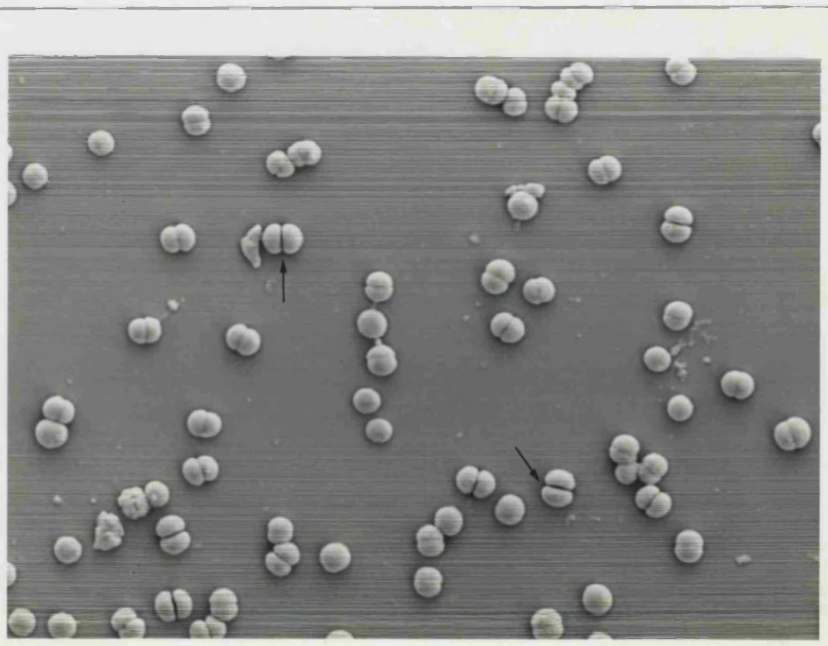
A



B



C



DAPI-stained nucleoids of wild type cells were relatively uniform in appearance when judged by intensity. Most (~80%) of the WT cells appeared not to be undergoing division. This backed up the results shown previously by scanning electron microscopy. Also there were only a few dead cells unstained with DAPI. A sample of DAPI stained WT *Synechocystis* cells is shown in Figure 3.2.4A.

By contrast DAPI-stained nucleoids in a sample of cells from ragged colonies were much more variable. Although some looked like wild type, approximately 50% of the cells stained less intensely. The latter included more than half of those in cell division (Fig. 3.2.4B). Although nucleoid segregation had occurred in many of these cells, in ~3% the DNA seemed to have mis-segregated.

DAPI-stained smooth mutants were also distinctive. About half of these cells (both dividing and non-dividing) carried normal amounts of DNA but the remainder had less. Occasional cells (~1%) had no DNA and were assumed to be dead. Mis-segregation was seen in about 1% of the cells. Another 1% of cells were larger than the rest and contained about twice as much as DNA, as judged by intensity, than wild type cells. An example of these is shown in Figure 3.2.4C.

3.2.4 Autofluorescence of *Synechocystis* transformants

The chlorophyll present in viable cells caused them to fluoresce orange whereas dead cells are green under UV light. When the autofluorescence of wild type and ragged mutant cells was compared (Iain Wilson) there were many fewer dead cells in the wild type samples (Fig. 3.2.5A) than in the ragged mutant samples (Fig. 3.2.5B). In the ragged mutants approximately 50% of the cells were found to be dead and of these approximately half appeared to be undergoing cytokinesis.

These results coupled with those obtained from both the DAPI staining and scanning electron microscopy led to the hypothesis that following the partial knockout of *ycf*24 ragged colonies might not after all be attributed to the death of homoplasmic cells. Alternatively, the phenotype was due to an increased rate of cell division causing gradual loss of DNA and subsequent cell death. This could also cause the mis-segregation of chromosomes containing the kanamycin-disrupted *ycf*24, resulting in daughter cells containing only the wild type form of *ycf*24. Cells lacking the disrupted gene would no longer be resistant to kanamycin and would therefore no longer be viable. To address the possibility of an abbreviated cell cycle time, WT and ragged mutants were re-examined by visual monitoring of individual cells.

Figure 3.2.4

DAPI stained wild type *Synechocystis* and *ycf*24 mutants.

A. Wild type. The nucleoids appear relatively uniform in appearance with very few dead (unstained) cells present.

B. Heteroplasmic ragged mutant clone T8. Some cells appear to contain as much DNA as WT (single arrow). However, many contain less as shown by the reduced intensity of staining (double arrow) and a few show mis-segregation of DNA (triple arrow). Many dead cells are also present.

C. Heteroplasmic smooth mutant clone SM8. Approximately half of the cells contain as much DNA as WT, the remainder contain less, with the exception of a small number of cells containing approximately double the amount (single arrow). Occasional dead and mis-segregated cells (double arrow) were also observed.

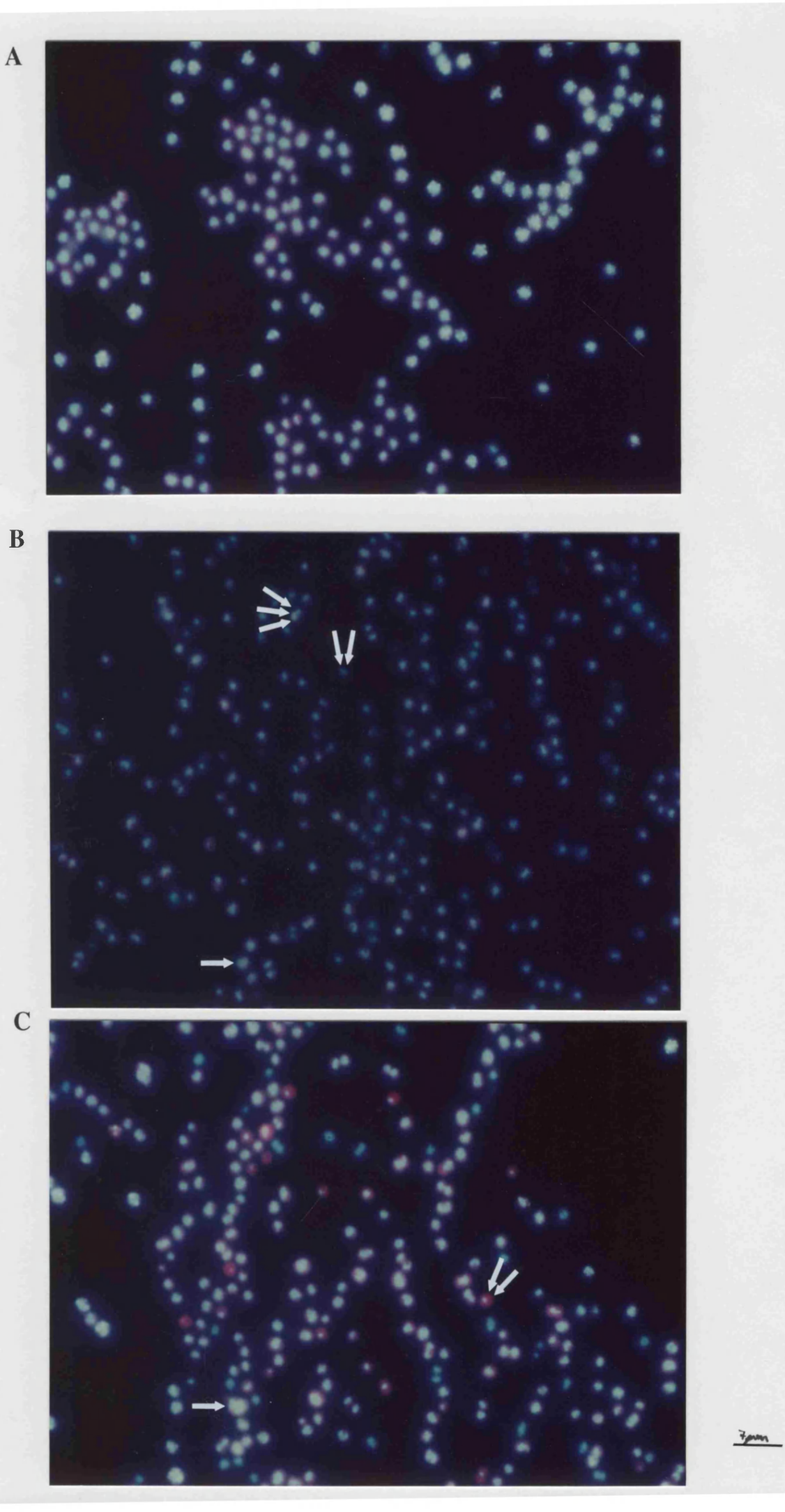

A
B
C

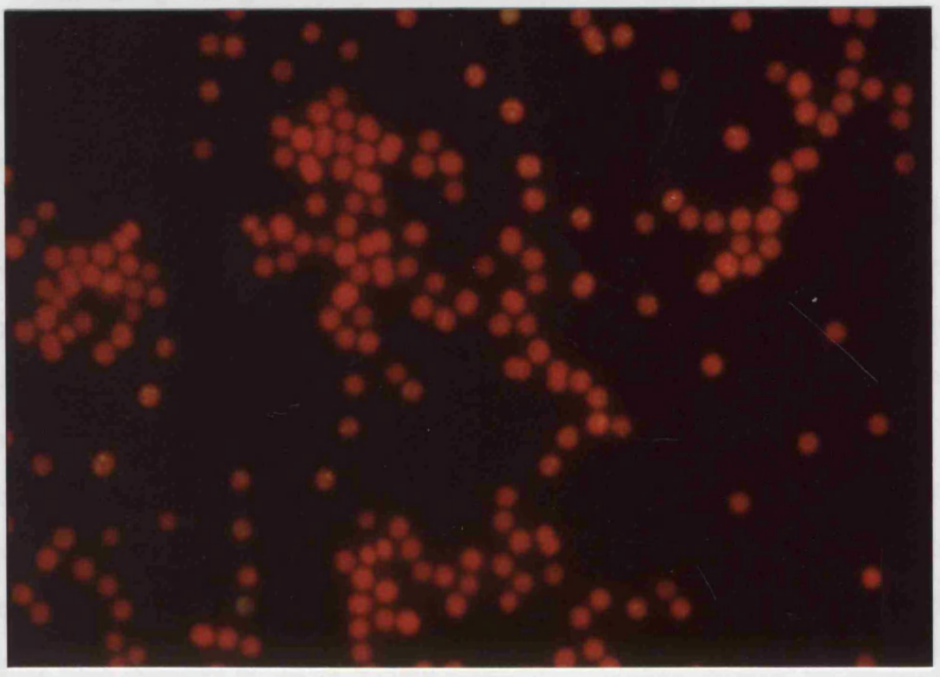
Figure 3.2.5

Auto-fluorescence of wild type *Synechocystis* and *ycf*24 mutants.

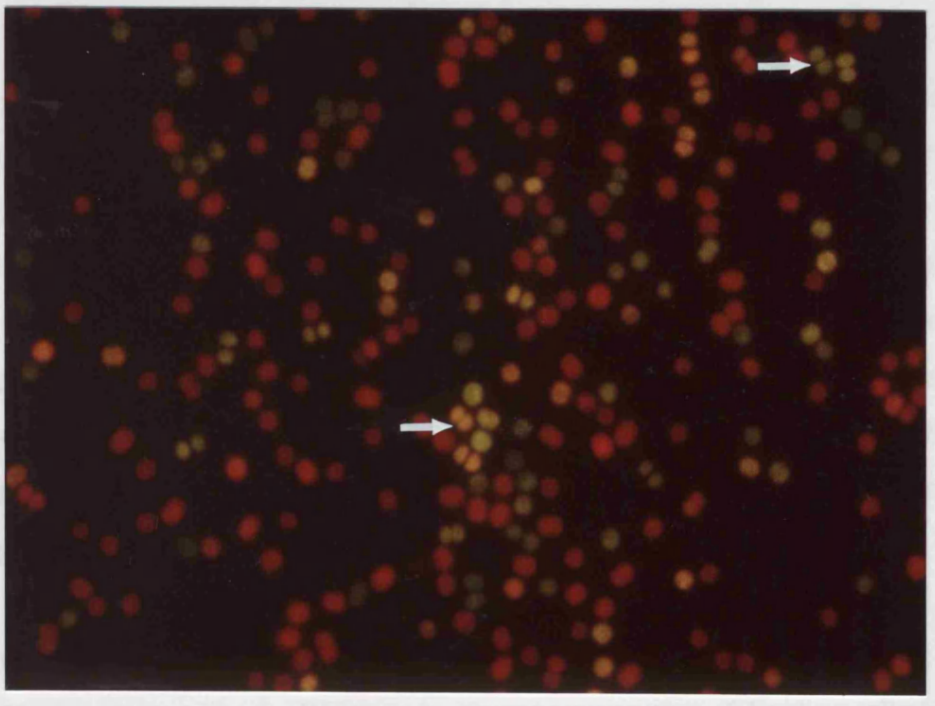
A. Wild type. The viable cells appear orange and the dead ones green. Few dead cells are present.

B. Heteroplasmic ragged mutant clone T8. Approximately half of the cells are dead. Half of these in turn appear to be in the late stages of cytokinesis as shown by the single arrows.

A



B



3.2.5 Growth rate of *Synechocystis* transformants

Estimating the growth rate *en masse* proved difficult because the ragged mutants could not be grown in liquid cultures nor could simple growth rates be measured by synchronisation by alternating light and dark cycles. This was due in part to the emergence of smooth mutants but also because the partial knockout caused a high incidence of cell death so that meaningful growth curves were not obtainable by simply measuring the optical density over a period of time. For the same reason, attempts to measure the growth rate of synchronous cultures by cell size using a coulter counter were abandoned. It seemed, therefore, the best way to answer the question of cell cycle length was to observe viable individual cells growing in a thin layer of BG11 agar over a period of a few generations. After various methods were explored, Conrad Mullineaux at UCL used a video-linked, phase contrast microscope to monitor the growth, over 5-6 generations, of a monolayer of cells on agar. The samples were returned to the 30°C light incubator between each time point.

Monitoring individual cells of wild type *Synechocystis* showed they had a doubling time of around eight hours and that they continued to divide indefinitely, producing smooth round colonies (Fig. 3.2.6). A similar doubling time was found for both the ragged and smooth mutants. More significantly, it was noted that cells of the ragged mutants often halted in the late stages of cytokinesis. Some cells divided once or twice successfully and then stopped, before dying. Thus non-viable cells continuously appeared in the samples. This would account for the large number (48%) of cells at the start of the experiment, already in the late stages of cell division, that did not grow any further (Table 3.2.3). Only 13.5% of cells from ragged colonies developed into colonies. Moreover, the failure of cytokinesis within the developing colonies resulted in an irregular perimeter (Fig. 3.2.6). A further observation of interest was that groups of about four daughter cells sometimes stopped together suggesting that cell death was "determined" a couple of generations ahead (Fig. 3.2.7).

Cells of the smooth mutant were more comparable to WT, with 37.5% of viable cells developing into round, smooth contoured colonies (Fig. 3.2.6; Table 3.2.3). However, unlike the WT sample, a greater proportion of dividing cells found at the start of the experiment did not grow further. One interpretation is that the compensating mutation is subject to segregation and hence, it is only carried by some

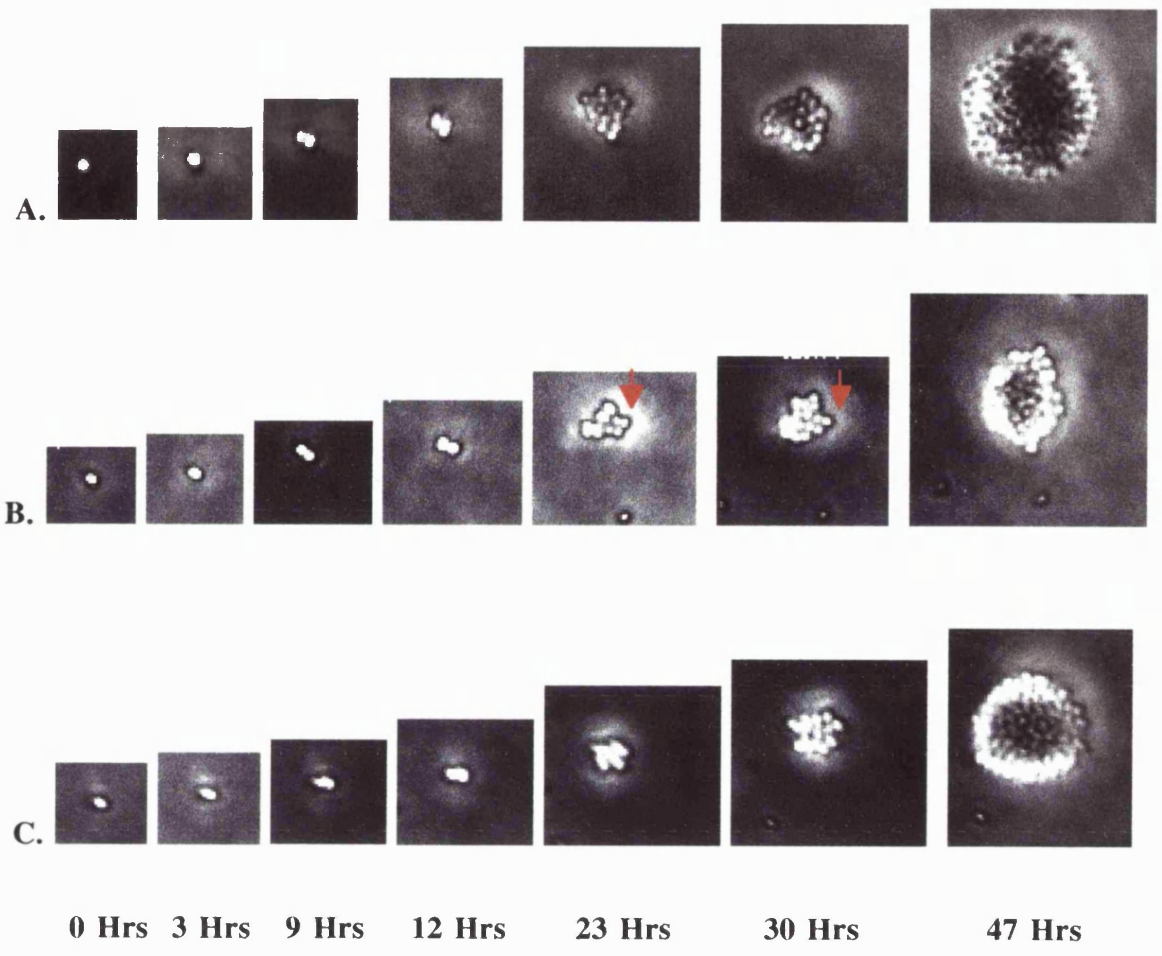


Figure 3.2.6

Colony development of *Synechocystis* PCC6803

A. Wild type. A single cell divides approximately once every 8 hours, forming a smooth, round colony.

B. T8 (heteroplasmic ragged mutant clone). Again division occurs every eight hours. However, the failure of some cells to divide within a developing colony (arrows) results in a ragged colony morphology.

C. SM8 (heteroplasmic smooth mutant clone). Normal division occurs resulting in a smooth, round colony.

	% Cells				
	Colony formers	Single cells (dead)	Double cells (dead)	Cells which divide then stop	Σ n
Wild Type	43	48	9	0	42
T8	13.5	23	48	15.5	52
SM8	37.5	37.5	25	0	16

Table 3.2.3

Growth of individual cells of WT *Synechocystis* and *ycf*24 mutants.

WT: Wild type *Synechocystis.*

T8: Heteroplasmic ragged mutant clone.

SM8: Heteroplasmic smooth mutant clone.

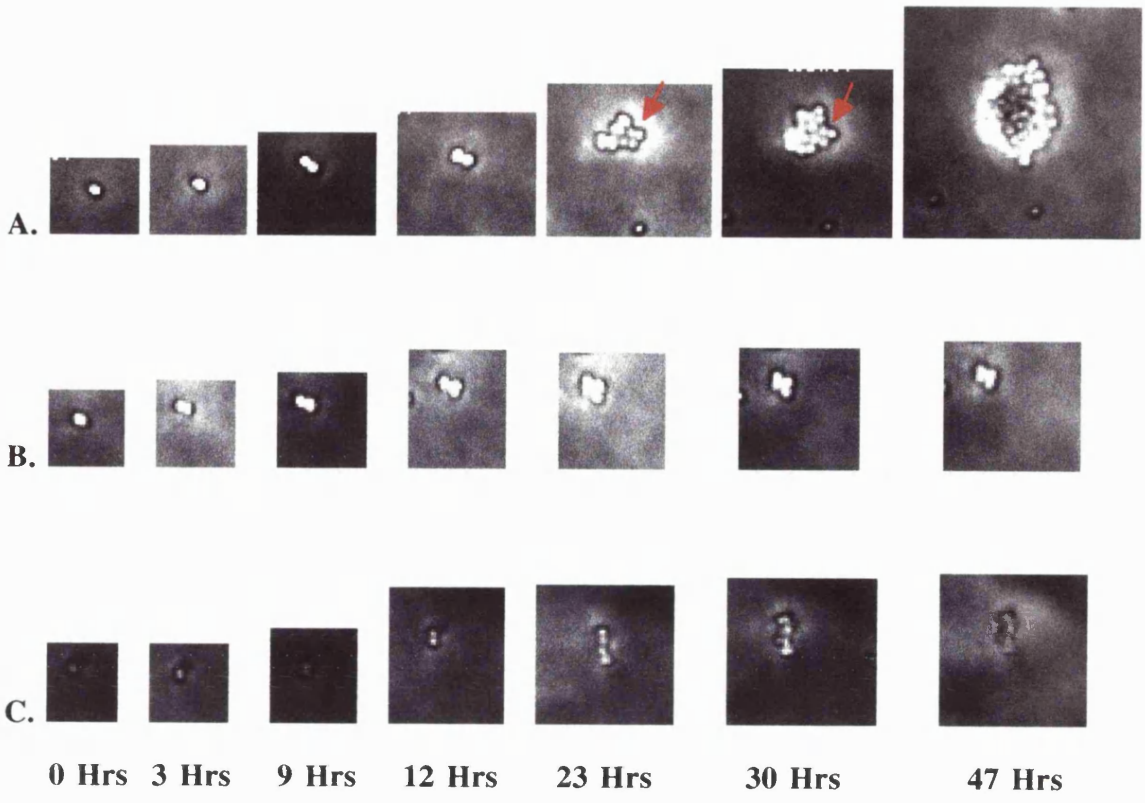


Figure 3.2.7

Cell division of ragged *ycf* 24 mutant clone (T8) of *Synechocystis* PCC6803

A. Ragged colony development. A single cell divides approximately once every 8 hours. However, the failure of some cells to divide within a developing colony (arrows) results in a ragged colony morphology.

B. A single cell divides once (12hrs) then twice (23hrs) into four cells before stopping.

C. A single cell divides once into two cells (12hrs). These then divide again giving four cells (23hrs). Division into 8 cells starts (30hrs) before cell death occurs.

of the cells. As a consequence, cells which carry this mutation no longer have defective cell division, and colonies develop normally.

3.2.6 Conclusion

In practise *Synechocystis* proved to be a more difficult organism to study than originally expected. Growth of the partial knockout of *ycf* 24 was very slow, making experiments extended in time and subject to contamination by fungi. The production of suppressor mutations when grown in liquid culture also made it difficult to produce enough material to isolate ragged mutant DNA for use in Southern blots. Usually many plates had to be harvested before making DNA. Southern blots for copy number analysis were also difficult due to problems found in completely digesting the DNA. Lastly, the incidence of compensatory mutations causing ragged colonies to become smoother increased after about a year of selection on kanamycin. Although DNA complementation experiments (Section 3.1.6) were carried out using relatively young mutants it is possible some of the observations were due to additional spontaneous mutations.

Although it was planned to do some over-expression work in *Synechocystis* and to determine how complementation might be occurring (see final Discussion), all the above factors made it difficult to draw conclusions. Accordingly, it was decided to examine *ycf* 24 knockouts in another surrogate organism, namely *E. coli* which contains only one copy of *ycf* 24 on its chromosome. The knockout of *E. coli ycf* 24, its over-expression and protein purification are covered in the following sections; Chapters 3.3, 3.4 and 3.5.

Despite the difficulties mentioned, analysis of the phenotype of individual "ragged" *ycf* 24 mutant cells confirmed that *ycf* 24 is likely to be an essential gene. Although cell cycle timing was maintained in these cells, mis-segregation and loss of DNA occurred and preceded death during cytokinesis. Furthermore, as some of the cells appeared to divide more than once or twice within a colony before stopping it is possible that Ycf 24 is required in stoichiometric amounts. In other words, if cells have enough copies of *ycf* 24, they are able to divide normally; with fewer copies cytokinesis is affected. This hypothesis, however, would not explain the observation of a small group of cells dying together (Fig 3.2.7C). Individual cells from smooth mutants also showed mis-segregation and varying DNA content. However, less cell death occurred during cytokinesis than in the ragged mutants, and many cells

developed into normal colonies. Only 25% of cells from SM8 were apparently stuck in cytokinesis. This contrasted with the results gained from SEM and suggests that cytokinesis in the cells with the smooth phenotype is dependent upon segregation of both the suppresser mutation as well as disrupted *ycf* 24. Conversely, cytokinesis in the ragged mutants is controlled by segregation of the disrupted gene only. However, perhaps not too much should be made of the smooth mutant observations as only a small number of cells was examined. Overall, these data suggest *ycf* 24 plays a role in cytokinesis but may also be involved in nucleoid division and/or replication.

3.3 Knockout of *E. coli ycf*24

3.3.1 Gene replacement using linear DNA

Karoui *et al.* (1999) found that gene replacements could be achieved in wild-type *E. coli*, with an efficiency of about 60 gene replacement events/μg of linear DNA, if electrocompetent cells were used. The presence of Chi sites to block linear DNA degradation were found to have no effect on efficiency and diminished RecBCD-mediated exonucleic activity was attributed to electrotransformation. Therefore this seemed a relatively simple way of performing the knockout of *E. coli ycf*24.

*E. coli ycf*24 was amplified (DNA from the Sure strain) using primers *ycf*24-5p (*Bam* HI) and *ycf* 24-3p (*Eco* RI), cloned directly into pCR 2.1 and used to transform *E. coli* strain INVαF'. Recombinants were selected by ampicillin resistance and blue/white colony colour. Following digestion with *Bsg* I, the *ycf* 24-pCR2.1 construct was disrupted by insertion of a streptomycin resistance cassette (Sp1). Recombinants were selected by growth on plates supplemented with 50μg/ml of streptomycin. Plasmid preparations were analysed by *Eco* RI, *Hind* III and *Xho* I restriction. Sequencing was also carried out using primers (A8 and A9) designed from sequence close to the inactivation site to confirm the correct cloning and the orientation of the *aadA* gene within *ycf*24. Due to the presence of both *Bam* HI and *Eco* RI sites within pCR 2.1 the *ycf*24 Sp1 insert was re-cloned into pBSKS+ and *ycf* 24 Sp1-pBSKS+ linearized with *Eco* RI. The linearized recombinant plasmid was then treated with alkaline phosphatase to prevent re-ligation and increase the probability of homologous recombination occurring between the insert and the *E. coli* chromosome. Linearized *ycf* 24 Sp1-pBSKS+ was used to transform the recombination proficient *E. coli* host strain LE292 by electroporation.

Any genome recombination event resulting in the replacement of the endogenous copy of *ycf*24 with the streptomycin disrupted gene and loss of vector sequence could be selected by the resistance and sensitivity to different antibiotics. Transformants resistant to streptomycin but sensitive to ampicillin would be assumed to be knockouts as streptomycin is carried within the disrupted gene whereas the gene for ampicillin resistance is carried on the vector, pBSKS+.

Transformants were initially selected by streptomycin resistance to ensure the presence of disrupted *ycf*24 and a few normal sized colonies were observed after 24

hours. However, when the clones were streaked onto plates supplemented with ampicillin they continued to grow. Resistance to both antibiotics indicated that either the vector had re-circularised or it had inserted into the genome as a tandem repeat alongside the endogenous copy of *ycf* 24. All of the clones were subsequently found by restriction digest to contain re-circularised vector. Normal sized colonies were also observed with the pUC18 (ampicillin selection) and circular *ycf* 24 Sp1-pBSKS+ (streptomycin selection) control transformations.

After 2 days incubation some very small colonies were also observed on the streptomycin plates. When these clones were re-streaked onto streptomycin they resulted in tiny colonies again. When a sample of these cells were fixed to a microscope slide and stained with Giemsa they were found to be elongated, unlike the pUC18 control transformed cells as shown in Figure 3.3.1. Figure 3.3.1A corresponds to the pUC18 control transformation and 3.3.1B to the linear *ycf* 24 Sp1-pBSKS+ transformed cells. The unusual appearance of the cells and colony size indicated that something had happened in these transformants. However, only resistance to streptomycin had been confirmed and for a knockout to be implicated the sensitivity of these clones to ampicillin needed to be determined. More colonies were picked but this time streaked onto both streptomycin and ampicillin plates. Colonies again were present after two days selection with streptomycin but they were tiny (barely visible with the naked eye and appeared more like a smear than discreet colonies). When observed under the microscope some colonies were round as usual, but most were elongated and mis-shapen which was attributed to cell lysis following cell death.

No colonies, bar one exception, grew up on ampicillin, suggesting that recombination had occurred but that the knockout was lethal. Amplification of DNA by PCR directly from colonies using primers (A20 and A21) designed to *E. coli* genomic sequence outside *ycf* 24 was attempted but was unsuccessful.

3.3.2 Complementation by over-expression of *ycf* 24

Because *ycf* 24 was found to be an essential gene in the cyanobacterium *Synechocystis* it was likely to be an essential gene in *E.coli.* In an attempt to complement the knockout in *E. coli* described above, I decided to co-transfect LE392 with linearized *ycf* 24 Sp1-pBSKS+ as well as an episome expressing WT *ycf* 24. Loss of viability of the knockout could then be recovered by the complementation.

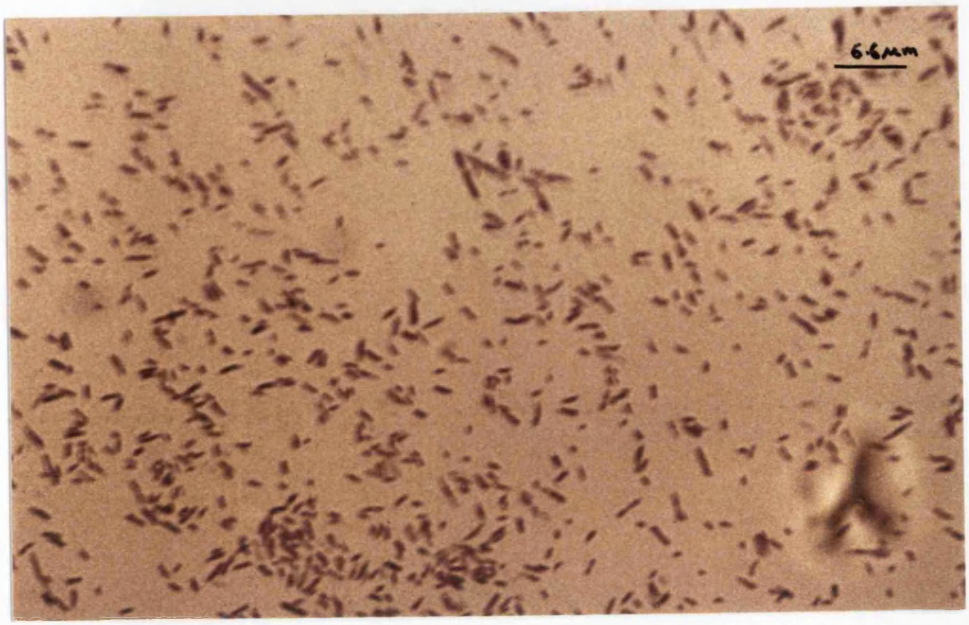
Figure 3.3.1

Cell morphology of LE392 cells following transformation with pUC18 and linear *ycf*24Sp1-pBSKS+.

A. pUC18 control. Cells appear to be normal sized.

B. Linear *ycf* 24Sp1-pBSKS+. Cells appear to be elongated, at least 3 times the length of the control cells.

A



B

Accordingly, *ycf* 24 was amplified using the primers pET28a-1 (*Nco* I) and pET28a-2 (*Xho* I), and cloned into the expression vector pET28. Recombinants were selected by kanamycin resistance and over-expression of *ycf* 24 as a C-terminal His-tagged protein was confirmed by induction with IPTG and western blotting using an anitbody against penta-His (see Chapter 3.5 and Figure 3.5.7).

The LE392 strain was transformed with *ycf* 24-pET-28a and selected on 50μg/ml kanamycin both in the presence and absence of 0.4mM IPTG. Normal sized colonies were observed as expected showing over-expression of *ycf* 24 in LE392 was not lethal and had no detrimental effect on colony size. Upon co-transfection with *ycf* 24Sp1-pBSKS+ (linearized plasmid) and *ycf* 24-PET-28a (circular expression vector) and selection on plates supplemented with 50μg/ml of both kanamycin and streptomycin no colonies were observed initially either in the presence of absence of IPTG. However, after two days selection small colonies were observed which when re-streaked grew very badly. Although some growth had occurred after two days of selection no distinct colonies were seen. When the plates were observed under the microscope, the faint streaky growth seemed to be comprised of thin "worm like" structures. This was unlikely to be contamination as a similar, if less severe, result was seen in the previous knockout experiment using linear *ycf* 24Sp1-pBSKS+ only. Also, no contamination was seen elsewhere on the plates or on streaked ampicillin plates. The growth phenotype was again attributed to cell death and lysis. Finally, as in the previous knockout experiment, no growth was apparent under ampicillin pressure suggesting pBSKS+ had been excised from the host genome. The results indicate that over expression of *ycf* 24 did not reverse lethality (i.e. complementation was unsuccessful). Because the *E. coli* knockout was still only implicated rather than confirmed, another method was chosen to try and demonstrate the knockout of *ycf* 24 in *E. coli*. The method chosen incorporates a number of elements to ensure the initial homologous recombination of the disrupted gene into the *E.coli* genome and a clearer selection of mutants, as described in the next section.

3.3.3 Knockout using the temperature sensitive suicide vector pKO3

The key elements of pKO3 include a temperature sensitive pSC101 (*repA*) origin of replication, the chloramphenicol resistance gene (*cat*) for positive selection

chloramphenicol acetyl transferase

and the *sacB* (levansucrase) gene for negative selection in 5% sucrose (Link *et al.* 1997).

ycf 24 Sp1 from pBSKS+ was amplified using the primers APK1 and APK2 and cloned into the *Bam* HI site of pKO3. Following transformation of *E. coli* strain INVαF′ by heat shock, transformants were selected at the permissive temperature, 30°C, on plates supplemented with 20 μg/ml of chloramphenicol and 50μg/ml of streptomycin.

Following analysis by restriction digestion and sequencing, one recombinant plasmid clone of *ycf* 24 Sp1-pKO3 was used to transform LE392. As the frequency of recombination tends to be low, 10^{-2} to 10^{-3} (Link *et al.*, 1997), transformants were initially selected for growth at 30°C on both chloramphenicol and streptomycin. This ensured all clones contained *ycf* 24Sp1 before selecting for integration of *ycf* 24Sp-pKO3 into the genome. Integrants were selected by serially diluting four colonies from the 30°C plates onto pre-warmed plates at 42°C. At this non permissive temperature, circular pKO3 episomes are unable to replicate due to the temperature sensitive origin of replication, *repA*. Consequently, strong expression of any antibiotic resistance genes carried by the vector (in this case *cat* and Sp1) can only occur if a single crossover event has taken place to incorporate the vector into the host genome. Expression of the genes carried on the vector are then under the control of the host. This temperature therefore selected for homologous recombination between the endogenous copy of *ycf* 24 and disrupted *ycf* 24 carried on pKO3, resulting in incorporation of *ycf* 24 Sp1-pKO3 into the host genome as an imperfect tandem duplication, and strong expression of Sp1 and *cat*. In practise the temperature sensitive origin of replication was not absolute, resulting in "slow growth" as opposed to "no growth" at 42°C. Therefore, to ensure clones were integrants only large colonies were used for the next stage of the experiment.

Four large colonies growing at 42°C were picked and serially diluted onto plates supplemented with 5% sucrose and 5% sucrose + streptomycin before incubating overnight at the permissive temperature of 30°C. As the expression of *sacB* carried on pKO3 is lethal to *E. coli* in the presence of sucrose, these growth conditions selected for a second crossover event (at point 1 or 2, Figure 3.3.2C), allowing replacement of the endogenous copy of *ycf* 24 and excision of the integrated plasmid. In theory, growth on sucrose only selected for replacement of the wild type

Figure 3.3.2

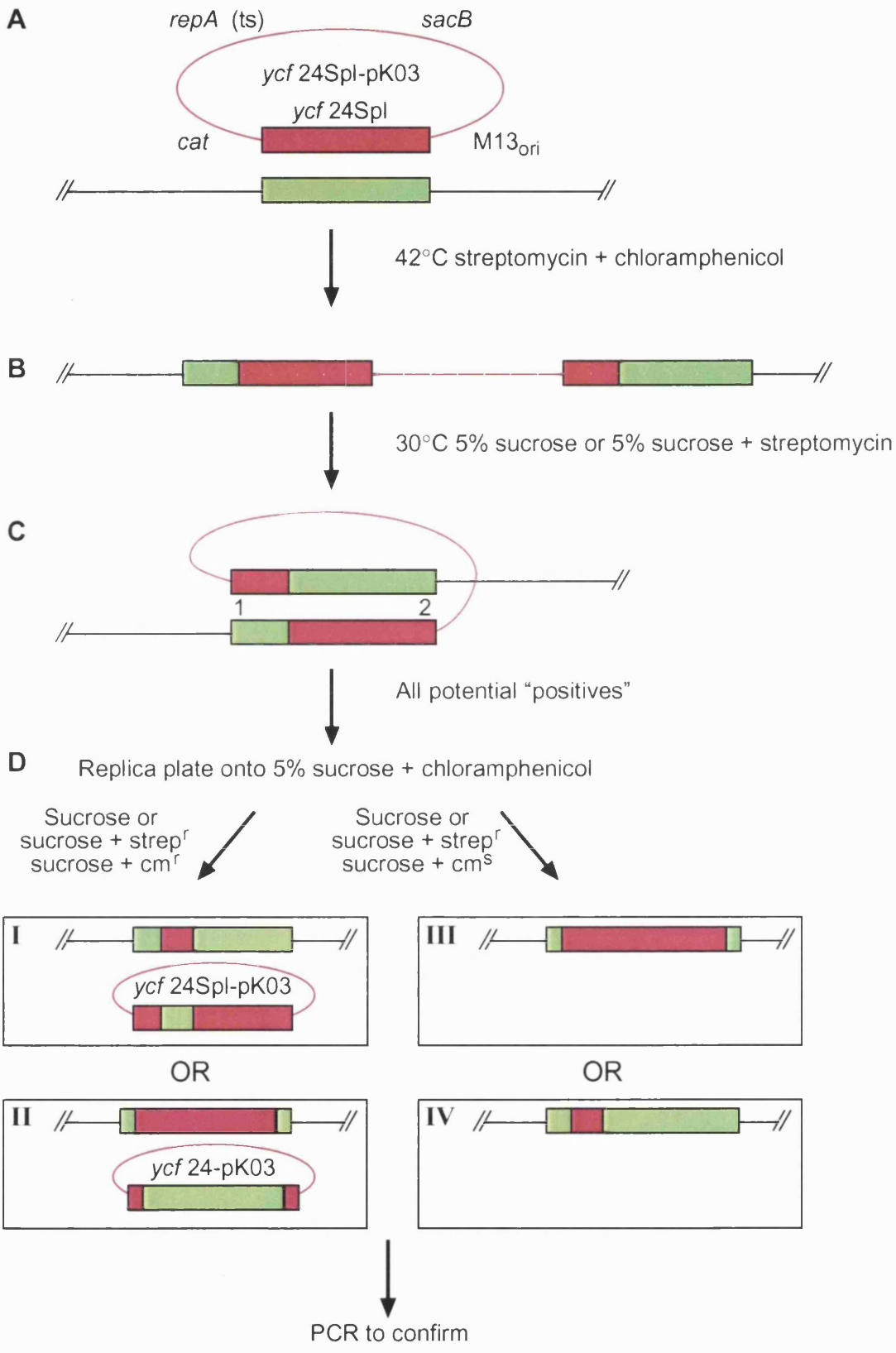
Knockout of *E. coli ycf* 24 using the temperature sensitive, gene replacement vector pKO3.
(Figure adapted from Link *et al.*, 1997)

A. LE392 transformed with *ycf* 24Sp1-pKO3 and selected on streptomycin and chloramphenicol at the permissive temperature, 30°C.
Result: Cells carry circular plasmid LE392 expressing Sp1 and *cat.* A small number of integrants are likely but not selected for.

B. Serial dilution of 30°C clones onto plates supplemented with both streptomycin and chloramphenicol at the non-permissive temperature, 42°C. As the circular episome is unable to replicate, streptomycin and chloramphenicol resistance is maintained by integrated plasmids.
Result: Single crossover event resulting in the integration of *ycf* 24Sp1-pKO3 at the *ycf* 24 locus (expression of Sp1 and *cat* under host control). Because the replication control of *repA* is not absolute, clones with integrated plasmid are larger than those without.

C. Serial dilution of large 42°C colonies onto 5% sucrose and 5% sucrose + streptomycin at 30°C.
Result: Expression of *sacB* renders pKO3 lethal in the presence of 5% sucrose. Growth on sucrose therefore selects for a second crossover event (at point 1 or 2) which results in further recombination at the *ycf* 24 locus and excision/loss of pKO3. Clones resistant to sucrose are potential positives and resistance to sucrose and streptomycin selects specifically for recombination of *ycf* 24Sp1. However, recombination resulting in replacement of *ycf* 24 with *ycf* 24Sp1 at the *ycf* 24 locus can not be confirmed until the clones are screened for their sensitivity to chloramphenicol.

D. Potential positives are replica plated onto plates supplemented with 5% sucrose and chloramphenicol because mutations in *sacB* would allow retention of circular plasmid in the presence of sucrose following excision from the *E. coli* genome.
Result: Clones resistant to sucrose + chloramphenicol (cm^r) contain circular plasmid. However, replacement of *ycf* 24 with *ycf* 24Sp1 may have occurred. The two possible outcomes (shown in I and II) can be distinguished by PCR.
Clones sensitive to sucrose + chloramphenicol (cm^s) are potential knockouts.
Clones found to be resistant to sucrose but sensitive to sucrose + chloramphenicol should contain either outcome III or IV (PCR will confirm which). Resistance to sucrose + streptomycin and sensitivity to sucrose + chloramphenicol however, selects for knockouts (homologous recombination at the *ycf* 24 locus resulting in its replacement with *ycf* 24 Sp1) as shown in III. PCR must be carried out to confirm this, as recombination of Sp1 can also occur at non-homologous sites.


A
repA (ts)
sacB
ycf 24Spl-pK03
ycf 24Spl
cat
M13ori
42°C streptomycin + chloramphenicol
B
30°C 5% sucrose or 5% sucrose + streptomycin
C
1
2
All potential "positives"
D
Replica plate onto 5% sucrose + chloramphenicol
Sucrose or
sucrose + strepr
sucrose + cmr
Sucrose or
sucrose + strepr
sucrose + cms
I
ycf 24Spl-pK03
III
OR
OR
II
ycf 24-pK03
IV
PCR to confirm

gene with either *ycf* 24 Sp1 or a recombinant wild type gene (III or IV, Figure 3.3.2D), whereas selection on sucrose and streptomycin sucrose selected for recombination of the disrupted gene only (III, Fig 3.3.2). In practise, before colonies could be screened for recombination and replacement of endogenous *ycf* 24 all colonies resistant to sucrose (i.e. those selected on 5% sucrose and 5% sucrose + streptomycin) had to be screened for their sensitivity to chloramphenicol. This was necessary to exclude any mutations that had occurred in *sacB* which would make pKO3 non-lethal in the presence of sucrose. Consequently all clones were replica plated onto 5% sucrose + chloramphenicol.

Two scenarios were observed with the four recombinant 42°C colonies used in this way. In two cases approximately 30 times more colonies were observed on the 5% sucrose plates than on the 5% sucrose + streptomycin plates. An example is shown in Table 3.3.1. All colonies observed on 5% sucrose were large, whereas on the 5% sucrose + streptomycin plates most of the colonies were small. In the remaining two experiments, colonies were only observed on the 5% sucrose plates and not on the 5% sucrose + streptomycin plates.

Replica plating showed that all colonies observed on the 5% sucrose + streptomycin plates also grew on 5% sucrose + chloramphenicol, indicating that pKO3 had been retained in the cells. However, as discussed by Link *et al.* (1997), it was still unknown whether the activity of *sacB* had been directy compromised by an internal mutation or whether a secondary mutation had taken place in the *E. coli* genome which conferred sucrose resistance. In these instances, where clones were resistant to sucrose + streptomycin and sucrose + chloramphenicol, one of two recombination events may have occurred. The first being recombination followed by excision of *ycf* 24 Sp1-pKO3 (I in Fig. 3.3.2D) and the second, excision of *ycf* 24-pKO3 following recombination of *ycf* 24 Sp1-pKO3 (II in Fig. 3.3.2D). Clones selected on 5% sucrose in the absence of streptomycin selection were rarely (1-2%) found to grow on 5% sucrose plates supplemented with chloramphenicol. This indicated that in the majority of cases mutations in *sacB* did not occur and, after the second recombination event, *ycf* 24Sp1-pKO3 was excised from the *E. coli* genome and lost (IV in Fig. 3.3.2D).

Approximately two hundred of these potential "positive" clones which grew on sucrose but not sucrose and chloramphenicol, were screened by PCR using primers

	Number of Colonies	
Dilution	5% sucrose	5% sucrose + streptomycin
1	300	10
2	24	1
3	6	-
4	2	-
5	1	-

Table 3.3.1

Number of colonies observed after serial dilution onto 5% sucrose and 5% sucrose + streptomycin. Data from *ycf* 24 knockout using DNA isolated from INVαF′ cells.

A20 and A21 designed from genomic DNA sequence either side of *ycf* 24. All PCR reactions resulted in a product the size of WT *ycf* 24 only (Figure 3.3.3). These results indicated that as no knockout had been achieved, *ycf* 24 is an essential gene in *E. coli*. It remained however, to prove that no recombination events had occurred.

Although the clones resistant to 5% sucrose + streptreptomycin and 5% sucrose + chloramphenicol were found to harbour pKO3, recombination events may still have occurred at the endogenous site of *ycf* 24 as mentioned above. Brown *et al.* (1995) showed that the essential gene, *murA*, could be replaced on the *E.coli* chromosome as long as another copy was present to complement it. However, PCR using primers designed to genomic sequence on either side of the recombination site amplified only the WT gene. Furthermore, PCR using primers either side of the cloning site within pKO3 (pKO3-L and pKO3-R) indicated the presence of *ycf* 24 Sp1-pKO3 (Fig. 3.3.4). This confirmed that replacement with the streptomycin resistance gene had not occurred at the *ycf* 24 locus (I as opposed to II in Fig. 3.3.2D).

The failure to "knock-out" *ycf* 24 as shown in the results so far, appeared to indicate that *ycf* 24 was an essential gene (personal communication, George Church, Harvard, USA). However, no recombination events had yet been observed directly. One possible explanation for the failure was that mis-matches had prevented integration, as the *ycf* 24 gene used for the knockout was amplified from the INVαF′ strain of *E. coli* and not that of LE392. To exclude this possibility the knockout experiment was repeated using LE392 isogenic DNA. A parallel knockout experiment was also undertaken as a positive control using the non-essential *E. coli* gene *ndh* (Calhoun *et al.* 1993).

3.3.4 Knockout using pKO3 and isogenic DNA

ycf 24 was amplified from purified LE392 DNA using primers APK1 and APK2. Following digestion with *Bam* HI it was cloned into pKO3 and selected at 30°C in the presence of chloramphenicol. Recombinants were screened for the *ycf* 24 insert by *Bam* HI restriction. SpI was cloned into the unique restriction site *Nru* I and clones containing *ycf* 24 Sp1-pKO3 sorted by resistance to streptomycin. Restriction analysis and sequencing confirmed the correct construct had been made.

A knockout experiment with this new construct was carried out, as described in the previous section. This time recombination of *ycf* 24 Sp1-pKO3 into the genome

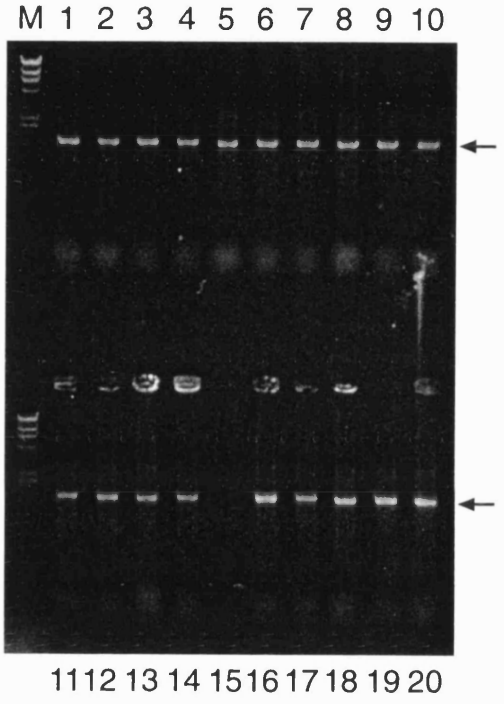


Figure 3.3.3

PCR of potential knockout clones from the *ycf* 24 knockout using DNA isolated from INVαF′ cells.

Lanes 1-14 and 16-20. 19 potential positive clones which grew on 5% sucrose but not 5% sucrose + chloramphenicol plates. All show a product of 1.5kb (arrows) corresponding to the size of endogenous *ycf* 24.

Lane 15. Negative control.

M. λHind III marker.

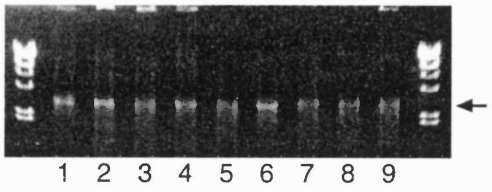


Figure 3.3.4

PCR of clones resistant to sucrose + streptomycin and sucrose + chloramphenicol using primers designed to vector sequence, either side of the *ycf* 24Sp1 cloning site.

Lanes 1-9. 9 clones found to grow on both streptomycin and chloramphenicol in the presence of 5% sucrose. All show a product of ~3kb corresponding to *ycf* 24Sp1.

was confirmed in four out of a sample of five large colonies grown at 42°C. This was done using primers spanning *ycf* 24 and it's two flanking genes (APK16 and APK17) in a long PCR reaction using LA Taq. Although no "long" products were obtained, in one reaction a product was observed which was the equivalent in size of WT *ycf* 24 and it's flanking genes i.e. no recombination having occurred (2.6kb as shown in Figure 3.3.5 lane 8). In the other samples, with the same reaction mix and approximate amounts of DNA, no products were amplified suggesting that recombination of the disrupted gene had occurred at the *ycf* 24 locus and prevented amplification of this product.

Six large colonies were selected at 42°C and serially diluted onto 5% sucrose and 5% sucrose + streptomycin plates. In all cases colonies were observed under both of these conditions but the number of colonies growing on streptomycin increased to approximately to one twentieth of those on sucrose only (Table 3.3.2). Again all colonies growing on 5% sucrose were large, whereas the colony sizes on plates containing 5% sucrose + streptomycin were much more variable. In addition to the occasional colony as large as those found on 5% sucrose (often only one or two on the plates with the most concentrated inoculum) most colonies were very small. Often these plates required two days growth at 30°C before they could be replica plated onto 5% sucrose + chloramphenicol. The very small colonies were assumed to be emerging *sacB* mutants as the majority were subsequently found to be resistant to chloramphenicol as well as streptomycin in the presence of 5% sucrose.

As described previously, all clones growing on both 5% sucrose and 5% sucrose + streptomycin were replica plated onto 5% sucrose + chloramphenicol to screen for mutations in *sacB* resulting in the retention of pKO3. Again clones selected on 5% sucrose only rarely grew up under chloramphenicol selection (i.e. the vector had been lost) and screening of 175 of these by PCR indicated the presence of WT *ycf* 24 and not the disrupted gene.

Unlike the previous knockouts experiment some clones (77) were found to grow in the presence of sucrose and streptomycin but not in the presence of sucrose and chloramphenicol. This indicated that *ycf* 24 had been replaced by *ycf* 24Sp1 (III in Fig. 3.3.2D) and *ycf* 24-pKO3 excised. This caused us to doubt that *ycf* 24 was an

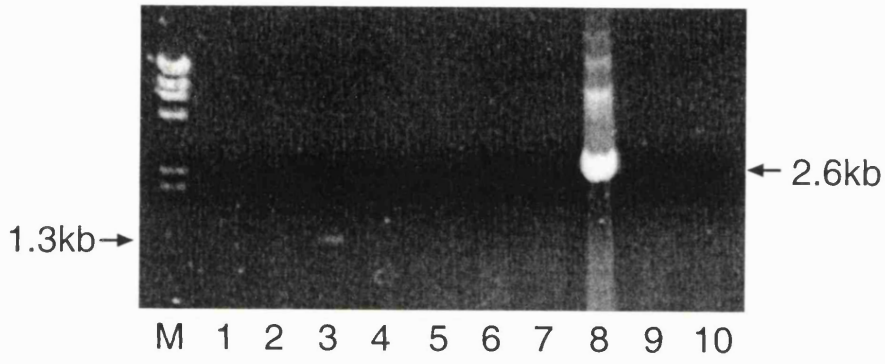


Figure 3.3.5

PCR to check for recombination of vector with insert sequence into the *E. coli* genome.

Lanes 1-5. *ndh* control knockout. A PCR product of approximately 1.3kb, the same size as endogenous *ndh*, is observed in lane 3 indicating that no recombination has occurred in this clone.

Lanes 6-10. *ycf* 24 knockout. A PCR product of approximately 2.6kb is marked in lane 8, the same size as *ycf* 24 and it's flanking genes. This again indicates that no recombination has occurred in this clone.

M. λHind III marker.

	Number of Colonies	
Dilution	5% sucrose	5% sucrose + streptomycin
1	1000	58
2	150	4
3	21	1
4	7	-
5	-	-

Table 3.3.2

Number of colonies observed after serial dilution onto 5% sucrose and 5% sucrose + streptomycin. Data from *ycf* 24 knockout using isogenic DNA isolated from LE392 cells.

essential gene in *E. coli* as originally indicated. However, PCR again showed that endogenous *ycf* 24 had not been replaced with the streptomycin disrupted version of *ycf* 24 (Fig. 3.3.6). Initially, to check that pKO3 had been excised and lost, minipreps were carried out for a sample of these clones resistant to sucrose + streptomycin but sensitive to sucrose + chloramphenicol. No plasmids were evident, confirming the loss of pKO3. The resistance of these clones to streptomycin indicated recombination of either the streptomycin resistance gene or *ycf* 24 Spl elsewhere on the *E. coli* chromosome. To confirm this, PCR was carried out with a sample of these clones using primers (APK12 and APK13) designed from sequence within *ycf* 24, either side of the streptomycin resistance gene. The resulting multiple products included bands of the expected size for the WT (~160bp) and streptomycin resistance genes (Fig. 3.3.7A). Although this initially indicated recombination of *ycf* 24 Spl had occurred elsewhere, multiple products of the same sizes were also observed using WT genomic DNA as a template in the same PCR reaction (Fig. 3.3.7B). This made it difficult to determine exactly what had happened.

As the PCR approach described above failed to determine if the streptomycin resistant potential knockout clones contained disrupted *ycf* 24Spl, Southern blots were carried out using probes for both *ycf* 24 and *aadA* (Spl). Genomic DNA was isolated from four clones resistant to 5% sucrose + streptomycin but sensitive to 5% sucrose + cam. Three (31, 32 and 35) originated from one 42°C clone, the fourth (50) originated from a second 42°C clone. Approximately 400ng of each DNA sample was digested with *Dra* III and 200ng of each was run in duplicate on an agarose gel. Following blotting, one half of the nitrocellulose membrane was hybridized by WT *ycf* 24 and the other to the radio-labelled streptomycin resistance gene (Spl). As *Dra* III cuts once within *ycf* 24 at position 1250 only two restriction products of approximately 4.8 and 2.2kb were expected in a Southern blot using WT *ycf* 24 as a probe. After hybridization with *ycf* 24, these were the only bands observed in all four clones (Fig. 3.3.8A). This indicated that following recombination, selection had favoured those bacteria carrying the streptomycin resistance gene recombined into the genome elsewhere than in the *ycf* 24 locus. Although it is not known where this recombination occurred, the Southern using Spl as a probe confirmed that the streptomycin resistance gene had recombined into the genome. Because as many as

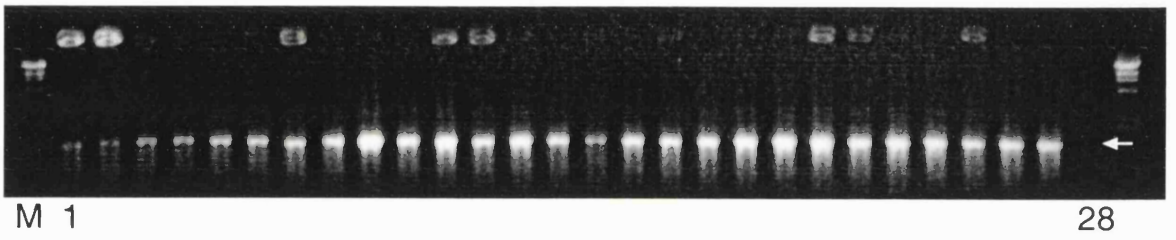


Figure 3.3.6

PCR of potential positive clones from the *ycf* 24 knockout using DNA isolated from LE392.

Lanes 1-27. Clones found to be resistant to 5% sucrose + streptomycin but sensitive to 5% sucrose + chloramphenicol. All show PCR products of 1.5kb (arrow), the same size as *ycf* 24 indicating that no knockout had survived.

Lane 28. Negative control.

M. λHind III marker.

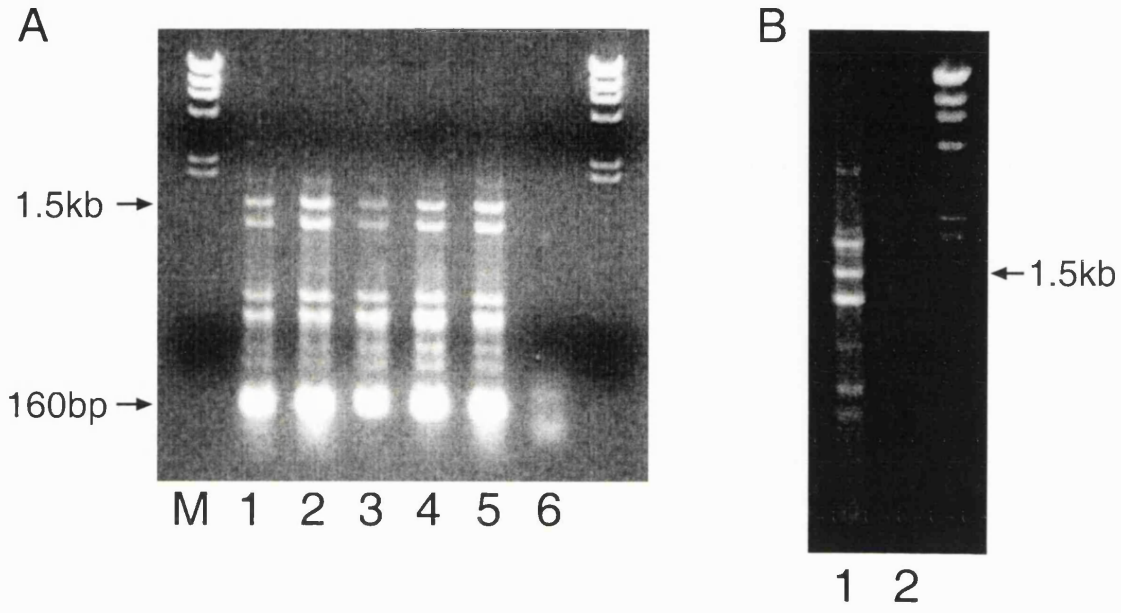


Figure 3.3.7

PCR of potential knockout clones in which (it turned out) no knockout had survived, using primers to *ycf* 24 sequence either side of the streptomycin resistance gene.

A. Lanes 1-5. Clones found to be resistant to streptomycin but sensitive to chloramphenicol in the presence of 5% sucrose. Multiple PCR products were observed, including one of around 160bp, the expected size for the WT gene and another of ~1.5kb, the same size of Sp1.

Lane 6. Negative control.

B. Lane 1. PCR using LE392 genomic DNA. Multiple products are also observed.

Lane 2. Negative control.

M. λHind III marker.

Figure 3.3.8

Southern blot of 4 recombinant clones resistant to streptomycin but sensitive to chloramphenicol (both in the presence of 5% sucrose) using probes for WT *ycf* 24 and Sp1.

A. Southern blot using WT *ycf*24 as a probe.

Lanes 1, 2 and 3. Clones originating from one large 42°C colony.

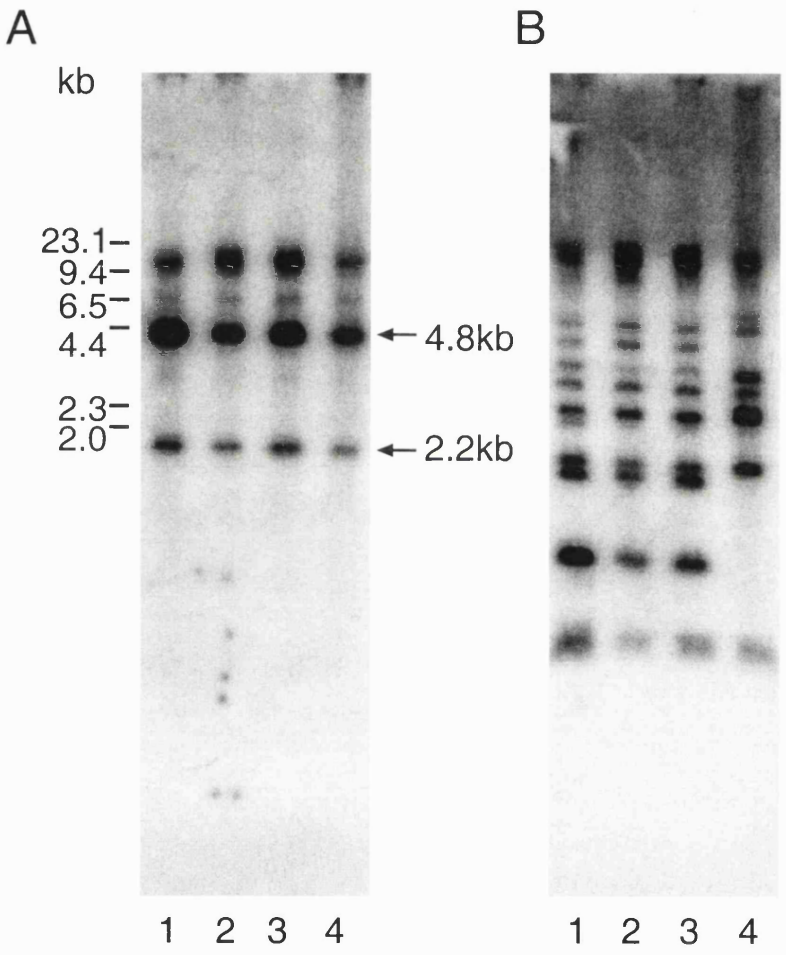
Lane 4. Clone originating from a second 42°C colony.

All four clones produce restriction products of approximately 4.8 and 2.2kb, the expected size for WT *ycf*24. No recombination elsewhere is indicated.

B. Southern blot using Sp1 as a probe.

Lanes 1, 2 and 3. Clones originating from one large 42°C colony. All three clones show the same restriction pattern indicating that the streptomycin resistance gene has recombined into the genome in a number of places.

Lane 4. Clone originating from a second 42°C colony. Again recombination of the streptomycin resistance gene has occurred. However the presence of additional and missing products indicates that recombination has occurred at different sites (as well as some of the same sites) to the 3 above clones.


A
B
kb
23.1
9.4
6.5
4.4
2.3
2.0
4.8kb
2.2kb
1 2 3 4
1 2 3 4

eleven bands were observed it seems likely that it had recombined into a number of places. Interestingly, the recombination sites seemed to vary from one 42°C clone to the next as expected from random recombination: clones 31, 32 and 35 gave the same pattern and clone 50 not only lacked some of the bands found in the previous three, but also had additional restriction fragments (Fig. 3.3.8B). These results not only proved that recombination had occurred at the non-permissive temperature but suggested that only recombination at non-homologous sites were recovered after selection, backing up the hypothesis that *ycf*24 is essential in *E. coli*.

In a final attempt to show whether homologous recombination at *ycf* 24 resulting in replacement of the endogenous gene with the streptomycin resistance gene had occurred, a sample of 100 clones resistant to sucrose, streptomycin and chloramphenicol were screened for knockout at the genomic site. As mentioned previously clones resistant to sucrose (or sucrose + streptomycin) and sucrose + chloramphenicol may have *ycf* 24Sp1 at the *ycf* 24 locus (see I and II, Fig 3.3.2D). However, as with the previous knockout, with non-isogenic DNA, no cases were observed in which the WT gene had replaced *ycf*24Sp1 in pKO3, allowing knockouts to remain viable.

3.3.5 Knockout of a non-essential *E. coli* gene – *ndh*

ndh codes for one of the two membrane-bound NADH dehydrogenases, *NDH*-2, involved in the catalysis of NADH:ubiquinol oxidoreductase activity and feeding of electrons into the quinone pool in the *E. coli* cytoplasmic membrane (Yagi, 1991). As it has been shown that the aerobic respiratory chain of *E. coli* can function with either *NDH*-1 or *NDH*-2 (Calhoun and Gennis, 1993; Calhoun *et al.* 1993) it was decided to replace this gene in *E. coli* with a disrupted form as a positive control for the pKO3 knockout.

ndh was amplified from a plasmid template provided by Nick Fisher (UCL), using primers APK6 and APK7. Following digestion with *Bam* HI it was then cloned into pKO3 and selected at 30°C in the presence of chloramphenicol. Recombinants were screened for the *ndh* insert by *Bam* HI restriction. SpI was cloned into the unique restriction site *Xmn* I and clones containing *ndh* Sp1-pKO3 were sorted by resistance to streptomycin and confirmed by restriction analysis and sequencing.

A knockout experiment was carried out as described previously with recombination of *ndh* Sp1-pKO3 confirmed by PCR in 4 out of 5 large colonies grown at 42°C. Again primers spanning *ndh* (APK14 and APK15) were used in long PCR using LA Taq and as before no long PCR products were obtained. However, a product was observed in one of the samples which was the equivalent in size to the WT *ndh* gene (1.3kb as shown in Figure 3.3.5, lane 3). This again suggests that recombination had occurred at the correct genomic site for the remaining four clones.

Six large colonies were selected at 42°C and serially diluted onto 5% sucrose or 5% sucrose + streptomycin as before. Growth was observed in all cases. This time the number of colonies growing on the 5% sucrose + streptomycin plates had increased to only one tenth of that on the sucrose plates (Table 3.3.3). As before colonies growing on 5% sucrose were large, whereas colonies on plates containing 5% sucrose + streptomycin were much more variable and required two days growth at 30°C before they could be replica plated onto 5% sucrose + chloramphenicol. Most of these were also resistant to chloramphenicol and when a sample of 150 clones was screened by PCR only the WT *ndh* gene was found. A sample of 50 of these is shown in Figure 3.3.9.

When the clones growing on 5% sucrose were replica plated onto 5% sucrose + chloramphenicol very few grew. However, as was the case with the *ycf* 24 knockout, none of the 160 "potential positive" clones screened by PCR were found to be *ndh* Sp1 knockouts. Only nine clones were found to be resistant to sucrose and streptomycin but sensitive to sucrose and chloramphenicol. Although they might have been *ndh* knockouts, none of them were found to have *ndh* Sp1 at the genomic site. As before, PCR resulted in products corresponding to the size of the endogenous copy of the gene (1.3kb). It is likely that the streptomycin resistance gene had again recombined into the *E. coli* genome at random, although this was not determined by Southern blots.

My results suggested that *ndh* was also an essential gene in *E. coli*, although it had been knocked out previously in *E.coli*, using a chloramphenicol resistance gene (Calhoun *et al.* 1993). This knockout was not done in the *E. coli* strain LE392 or using pKO3 so it is possible that in this strain, under the growth conditions used here, *ndh* is an essential gene.

	Number of Colonies	
Dilution	5% sucrose	5% sucrose + streptomycin
1	645	80
2	53	5
3	6	2
4	-	-
5	-	-

Table 3.3.3

Number of colonies observed after serial dilution onto 5% sucrose and 5% sucrose + streptomycin. Data from *ndh* knockout.

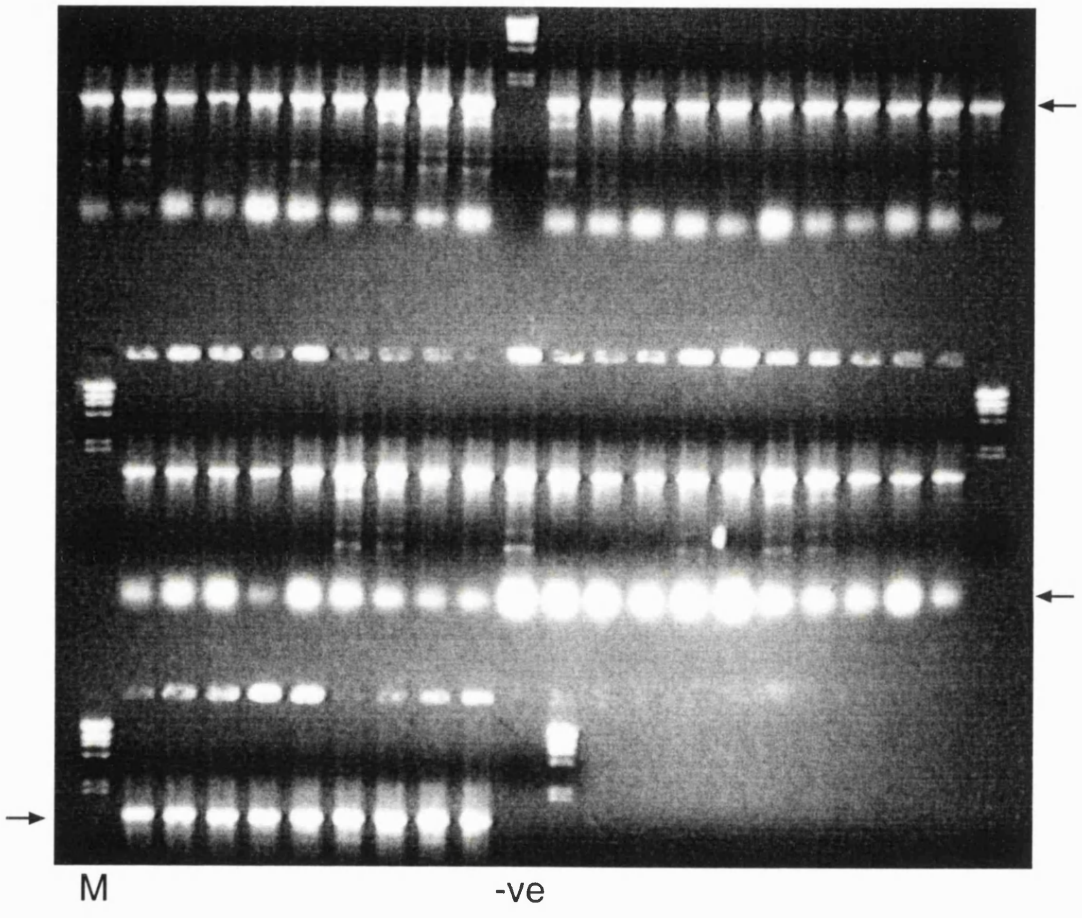


Figure 3.3.9

PCR of 50 clones resistant to both streptomycin and chloramphenicol in the presence of 5% sucrose, following the attempted knockout of *ndh*.

All show a product of ~1.3kb (arrows) corresponding to *ndh*. The blank lane corresponds to the negative control.

M. λHind III marker.

3.3.6 Conclusion

Data from both the *Synechocystis* and *E. coli* knockout experiments indicated that *ycf* 24 is an essential gene. Although the partial knockout in *Synechocystis* suggested a role in cytokinesis, the attempted knockout of *E. coli ycf* 24 shed no more light on it's function. The effects of its over-expression in *E. coli* are covered in Chapter 3.4.

3.4 Over-expression of *E. coli ycf*24

3.4.2 Over-expression of *ycf*24 in pMAL-c2

*E. coli ycf*24 was amplified by PCR from DNA isolated from INVα F'cells. The primers used, A11 and A9*, were designed to include an *Eco* RI and *Pst* I site respectively to allow cloning into the expression vector, pMAL-c2. Expression of coding sequence cloned into this vector results in the cytoplasmic expression of a maltose binding protein (MBP) fusion upon induction with IPTG. Following transformation by heat shock into *E. coli* SURE cells, transformants were detected by ampicillin resistance and blue/white colour selection. Analysis of clones by restriction with *Eco* RI and *Pst* I showed the majority of recombinant clones selected contained inserts of the expected size for *ycf*24 (~ 1.5kb). Initially four clones were selected and partially sequenced, using primers specific to vector sequence (MalE and M13; New England Biolabs), to check the sequence was in frame. As all appeared to be correct it was decided to induce all four clones.

500µl of overnight culture for each clone was used to inoculate 50ml of rich broth (L-broth + 0.2% D-glucose) supplemented with 100µg/ml of ampicillin and grown to an optical density at A_{600} of between 0.5 and 0.8. The 50ml cultures were then split into two flasks and one induced by the addition of IPTG to a final concentration of 0.3mM. Samples were taken from both the uninduced and induced flasks at time points t = 0, 1, 2 and 3 hours after the start of induction. 1ml samples were spun down and 100µl of SDS-PAGE sample buffer added. An additional 1ml sample was removed and the optical density measured as a decrease in the growth rate in induced cultures compared to the uninduced controls is often an indication of over-expression.

Little or no effect on the growth rate was observed when optical density readings were taken before and at one hour intervals after induction (Table 3.4.1) suggesting that any over-expression of a maltose binding protein fusion was not toxic. Western blots using an antibody against MBP, were carried out on protein samples collected for all of the clones induced. Detection of a product the size of MBP (48kDa) indicated that no fusion protein had been expressed from clone 1. However for clones 2, 3 and 4 induction of a fusion protein between the 83kDa and 175kDa markers was observed. Degradation of the fusion protein was apparent over

Optical density (A_{600})		t = 0h	t = 1h	t = 2h	t = 3h
Clone 1	U	0.535	0.890	1.309	1.445
	I		0.913	1.340	1.480
Clone 2	U	0.674	1.097	1.318	1.449
	I		1.088	1.307	1.436
Clone 3	U	0.783	1.180	1.364	1.497
	I		1.208	1.382	1.470
Clone 4	U	0.802	1.187	1.409	1.485
	I		1.250	1.471	1.530
INVαF' host	U	0.651	1.142	1.517	1.751
	I		1.141	1.515	1.730

Table 3.4.1

Optical densities (A_{600}) of uninduced (U) and induced (I) cultures containing clones 1-4 of recombinant *ycf*24-pMAL-c2 plasmids or no plasmid.

time, as was some leaky expression from the pMAL promoter in the uninduced cells. A western blot showing induction of clone 2 is pictured in Figure 3.4.1 and a Coomassie stained gel in Figure 3.4.2. All of the clones mentioned were subsequently sequenced using various primers including a number of primers internal to *ycf* 24 (see Appendix 3). As the sequences were found to be correct, and the expected size for the fusion protein was approximately 100kDa, the fusion proteins produced from all four clones were assumed to be full length.

Although the results given previously suggested that cytoplasmic expression of MBP-Ycf 24 was not detrimental to the host, impairment of growth was seen with a reduction of cell density before induction. Examples of three induction experiments using clone 2 (*ycf* 24-pMAL-c2) are shown in Table 3.4.2, (t = 0h and t = 3h time points only). Also included are the measurements collected after over-expression of the pMAL-c2 vector only control. Some reduction in the growth rate was also observed in these cells.

During the above experiment additional samples were taken for morphological examination. Samples of 1ml for the INVαF′ host controls, and *ycf* 24-pMAL-c2 containing cells before and after induction were taken at each time point, washed and then resuspended in 1ml of 1 x PBS. From each, 2µl was spotted onto a microscope slide (already prepared with poly-L-lysine), fixed and stained with Giemsa's stain. Interestingly, filamentation was observed following over-production of the MBP-Ycf 24 fusion. This was not observed in the uninduced and host cell controls. Filamenting cells ranged from 2 to 4 times the length of both the uninduced and host cell controls as shown in Figure 3.4.3.

3.4.2 Over-expression of MBP-Ycf24 in KM4104

To determine whether filamentation was observed in a different cell line, *ycf* 24-pMAL-c2 was cloned into the Δ*recA* strain, KM4104. The use of a cell line which is unable to produce functional RecA was important as it prevents induction of the fifteen or so SOS response genes. One of these is *sulA* whose protein product acts to inhibit cell division (Waller, 1996).

KM4104 was transformed with 1µl of a miniprep of clone 2 (*ycf* 24-pMAl-c2) by heat shock and transformants selected by ampicillin resistance. Four clones were checked for the presence of *ycf* 24-pMAL-c2. Following induction with 0.3mM IPTG,

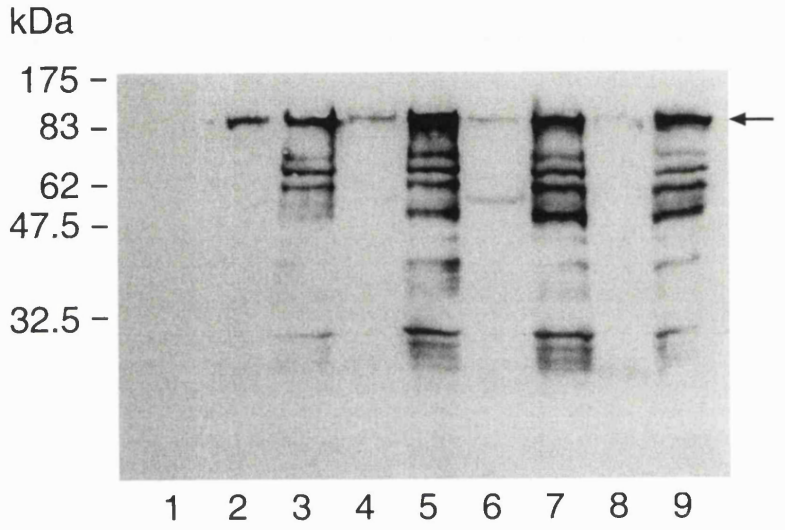


Figure 3.4.1

Western blot of full length fusion protein following induction of *ycf* 24-pMAL-c2.

Lane: 1. t = 0h.
2. t = 1hU.
3. t = 1hI.
4. t = 2hU.
5. t = 2hI.
6. t = 3hU.
7. t = 3hI.
8. t = 4hU.
9. t = 4hI.

A band around 100kDa (arrow) shows expression of full length MBP-Ycf 24.

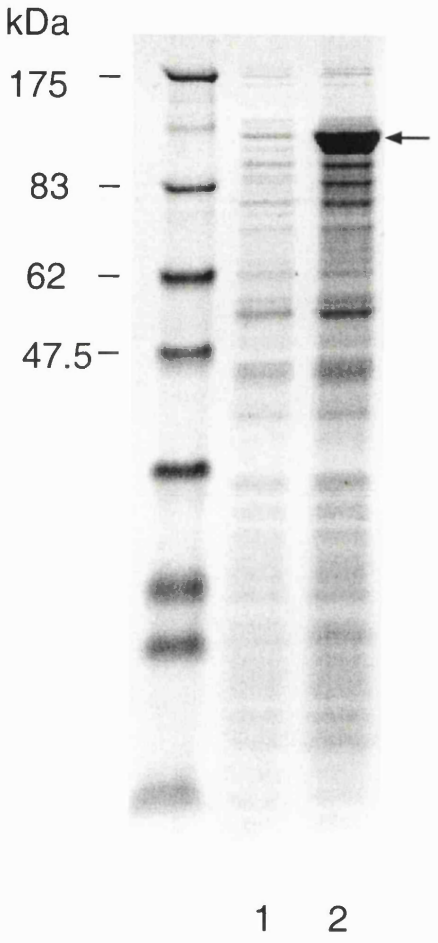


Figure 3.4.2

Coomassie stained SDS-PAGE gel following induction of full length MBP-Ycf 24.

Lane: 1. t = 0h.
2. t = 3hI.

A band around 100kDa (arrow) again indicates expression of the full length fusion protein from *ycf* 24-pMAL-c2.

Optical density (A_{600})		Experiment 1		Experiment 2		Experiment 3	
		t = 0h	t = 3h	t = 0h	t = 3h	t = 0h	t = 3h
INVαF′ host	U	0.253	1.348	0.490	1.346	0.307	1.325
	I		1.333		1.376		1.356
Vector	U	0.243	1.147	0.274	1.424	0.270	1.206
	I		1.145		1.240		1.097
Clone 2	U	0.272	1.156	0.385	1.344	0.236	1.134
	I		1.034		0.940		1.035

Table 3.4.2

Optical densities (A_{600}) of uninduced (U) and induced (I) cultures containing clone 2 of recombinant *ycf*24-pMAL-c2 plasmid, pMAL-c2 only or no plasmid.

Figure 3.4.3

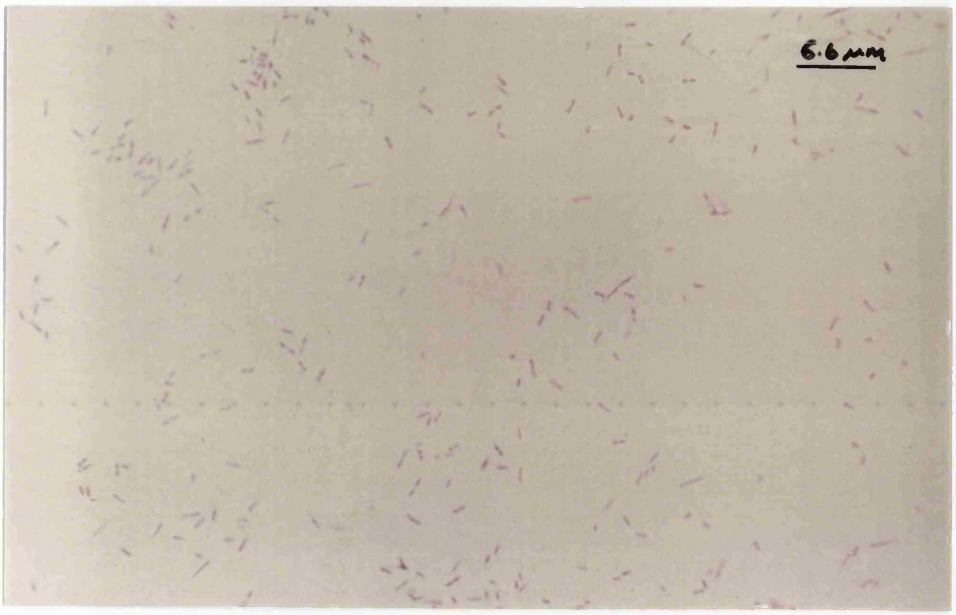
Giemsa stained INVαF' cells following induction of *ycf*24-pMAL-c2.

A. Uninduced cells.

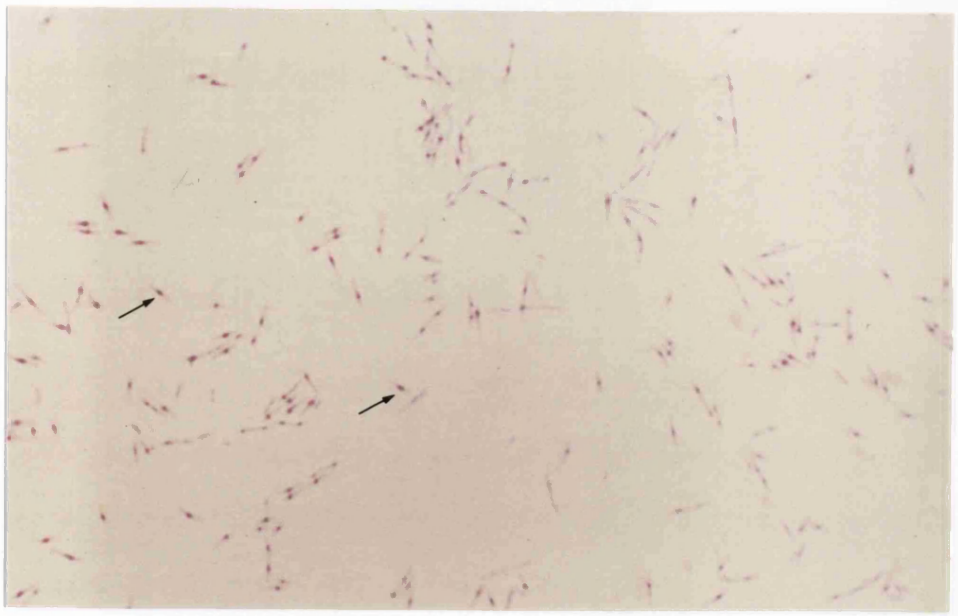
B. Cells following 3hrs of induction with IPTG. Cells appear elongated, approximately 2-4x the length of the uninduced cells.

The arrows indicate areas of intense staining causing cell distortion. These were presumed to be the nucleoids (see Section 3.4.3).

A



B



growth rates were measured by optical density A_{600}, before and after induction. Induced KM4104 cultures showed a higher growth rate initially, followed by a slight decrease when compared to the uninduced cultures (Table 3.4.3). Induction of full length fusion protein (~100kDa) was confirmed by western blots for all clones. One example (clone 8) is shown in Figure 3.4.4. Again some degradation was apparent, as well as leaky expression in the uninduced cultures.

When cells were checked for filamentation, it was observed that they were at least three times the length of the uninduced cells and host controls. In fact filamentation appeared to be more severe than in INVαF′ cells over-expressing *ycf* 24-pMAL-c2. Examples of these cells are shown in Figure 3.4.5.

When INVαF′ cells containing pMAL-c2 alone (MBP-αLacZ fusion) were over-expressed and stained with Giemsa's stain they also were found to have moderate filamentation as shown in Figure 3.4.6. This indicated that the phenotype was due to over-expression of the MBP tag as opposed to over-expression of the Ycf 24 fusion protein. This realisation caused me to go back and reconsider the cells expressing *ycf* 24 as a maltose binding protein fusion to see if any other affects were apparent.

3.4.3 Analysis of nucleoids by DAPI staining

During the analysis of Giemsa-stained KM4104 and INVαF′ cells expressing MBP-Ycf 24 it was observed that the cells appeared distorted and thickened around areas of intense staining (see arrows in Figs. 3.4.3B and 3.4.5B). These structures tended to occur once or twice along the length of the elongated cells, most often in the centre. Filamenting cells in the controls tended to have smaller, multiple areas of dark staining within the cells. This included the filamenting cells following induction of pMAL-c2 as shown in Figure 3.4.6B. These structures were assumed to be nucleoids and led to the investigation of nucleoids, by DAPI staining, in cells over-expressing MBP fusion proteins.

Cell samples were fixed onto microscope slides and incubated with DAPI before observation under UV light microscopy. Nucleoids were analysed in both KM4104 and INVαF′ host cells expressing full length MBP-Ycf 24 fusion protein. In both cases DAPI staining highlighted the internal structures shown by Giemsa. These were single or double areas of very bright staining of DNA within each elongated cell

Optical density (A_{600})		t = 0h	t = 1h	t = 2h	t = 3h
Clone 8	U	0.546	1.071	1.451	1.587
	I		1.113	1.380	1.443
Clone 10	U	0.500	1.002	1.424	1.614
	I		0.990	1.314	1.387
Clone 11	U	0.553	1.044	1.468	1.612
	I		1.033	1.304	1.392
Clone 12	U	0.543	1.128	1.455	1.542
	I		1.162	1.276	1.394
Host KM4104	U	0.630	1.142	1.481	1.603
	I		1.141	1.494	1.595

Table 3.4.3

Optical densities (A_{600}) of uninduced (U) and induced (I) KM4104 cultures containing recombinant *ycf*24-pMAL-c2 plasmids or no plasmid.

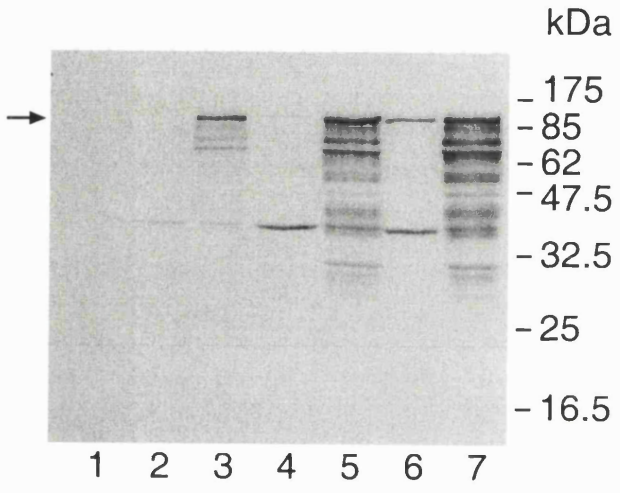


Figure 3.4.4

Western blot of full length MBP-Ycf 24 in KM4104 host cells.

Lane: 1. t = 0h.
2. t = 1hU.
3. t = 1hI.
4. t = 2hU.
5. t = 2hI.
6. t = 3hU.
7. t = 3hI.

As with expression in INVαF′ cells expression of a fusion protein of approximately 100kDa (arrow) is indicated.

Figure 3.4.5

Giemsa stained KM4104 cells following over-expression of MBP-Ycf 24.

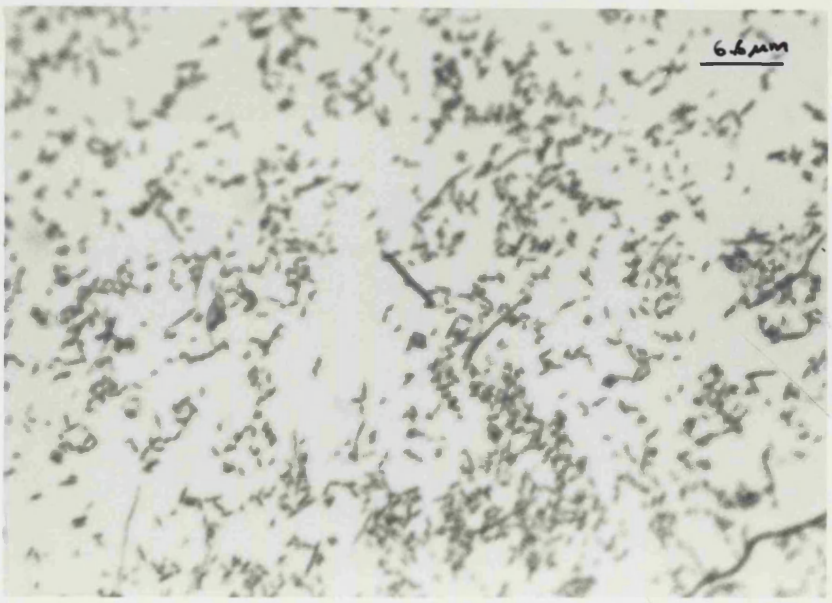
A. Uninduced host cells.

B. Uninduced cells containing *ycf* 24-pMAL-c2 (t = 3h).

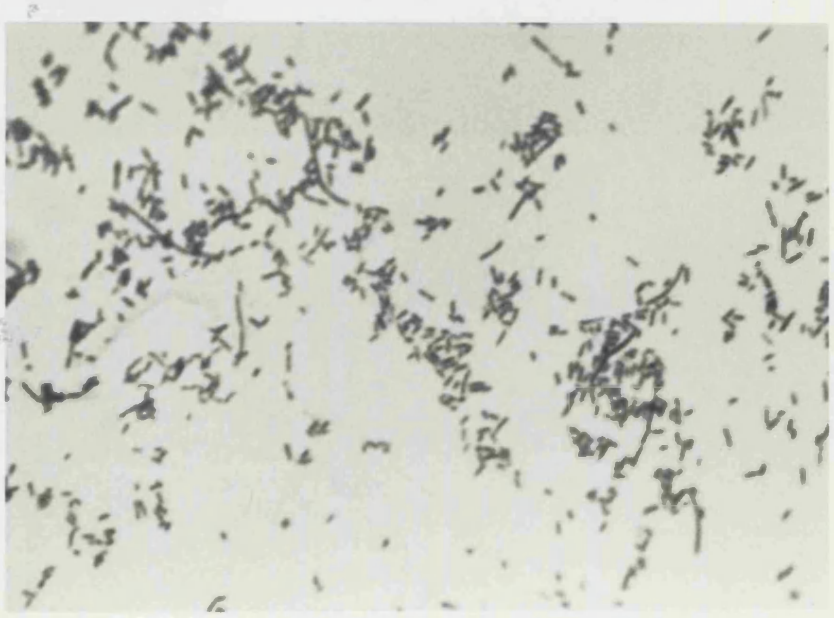
C. Induced cells containing *ycf* 24-pMAL-c2 (t = 3h).

Filamentation is apparent in the cells over-expressing MBP-Ycf 24. This filamentation effect appears more severe than in INVαF′ cells.
The arrows indicate areas of intense staining and cell distortion. These were presumed to be the nucleoids (see Section 3.4.3).

A



B



C

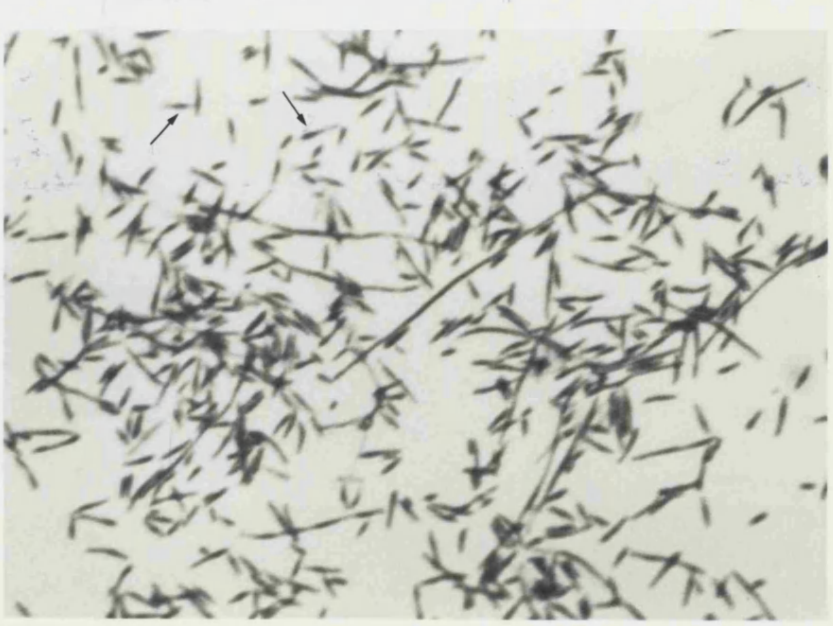


Figure 3.4.6

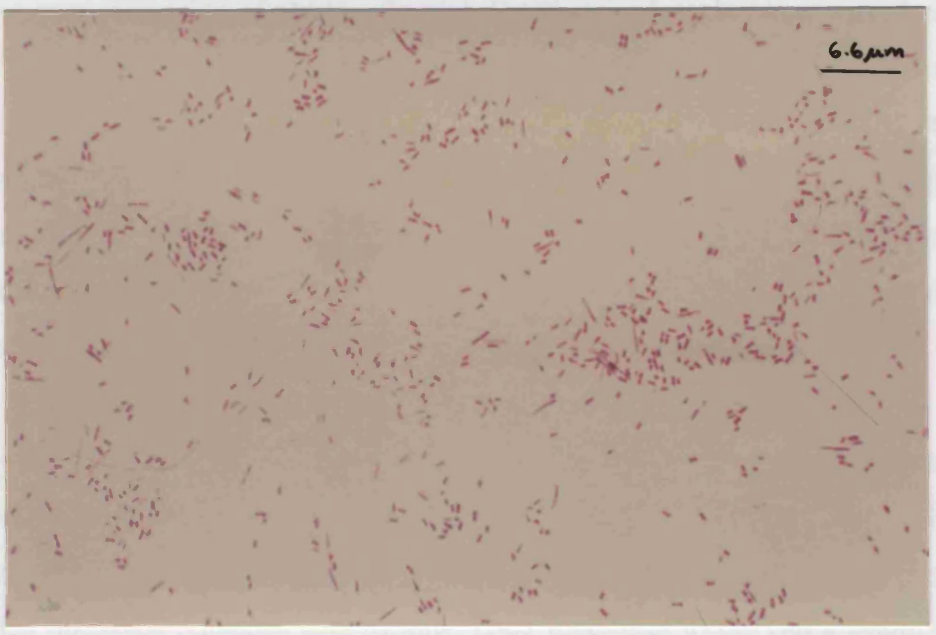
Giemsa stained INVαF' cells following induction of pMAL-c2.

A. Uninduced cells (t = 3h).

B. Induced cells (t = 3h). Moderate filamentation is observed as with over–expression of the Ycf 24 fusion protein.

The arrows indicate areas of intense staining. These were presumed to be the nucleoids (see Section 3.4.3).

A



B

(Fig. 3.4.7). In cells expressing the MBP-LacZα fusion protein of ~52kDa (produced by the over-expression of pMAL-c2 only), DAPI stained nucleoids appeared normal despite moderate filamentation. In these cells multiple nucleoids were found to span the length of elongated cells. This was also seen in elongated cells occurring occasionally in the host control cultures. Sometimes the nucleoids were regularly spaced along the entire length of the elongated cell, at other times clusters of nucleoids running parallel to one another were observed at either end and/or central to the cell (arrows in Fig. 3.4.8).

The enlarged size of the nucleoids, causing cell thickening in the cells over-expressing the Ycf 24 fusion protein, suggested that DNA replication and nucleoid formation had continued. It appeared therefore, that the nucleoids were condensed because of their failure to segregate after division. Closer examination showed divided nucleoids remained side by side, lying transverse to the cell's length rather than parallel to it in the usual way. Dumbbell shaped nucleoids suggesting this are shown by a single arrow in Figure 3.4.9. The next nuclear division was found to be longitudinal but the products remained clustered giving the appearance of double dumbbell structures which are indicated by a double arrow, also shown in Figure 3.4.9.

To test whether this phenotype was restricted to the over-expression of MBP-Ycf 24, a fragment coding for 340 amino acid residues (581 to 921) of the merozoite surface protein I from *Plasmodium chabaudi chabaudi* was isolated by *Eco* RI digestion from a recombinant pMAL-p2 clone, pMCK113 (provided by Stuart Quinn, NIMR). This similar sized insert to *ycf* 24 was cloned into pMAL-c2 and used to transform *E. coli* strain INVαF'. Over-expression proved to be lethal after three hours. However samples were still taken and observed for changes in morphology. Moderate filamentation was observed as before, however the DAPI stained nucleoids of viable cells appeared normal (Fig. 3.4.10). This indicated that nucleoid condensation was a direct result of over-expression of *ycf* 24 as an MBP fusion protein.

3.4.4 Over-expression of truncated *ycf* 24 fusions

To investigate which region of *ycf* 24 caused the nucleoids to mis-segregate a 504bp 3' truncated portion of *ycf* 24 was amplified from *E.coli* using the primers A11

Figure 3.4.7

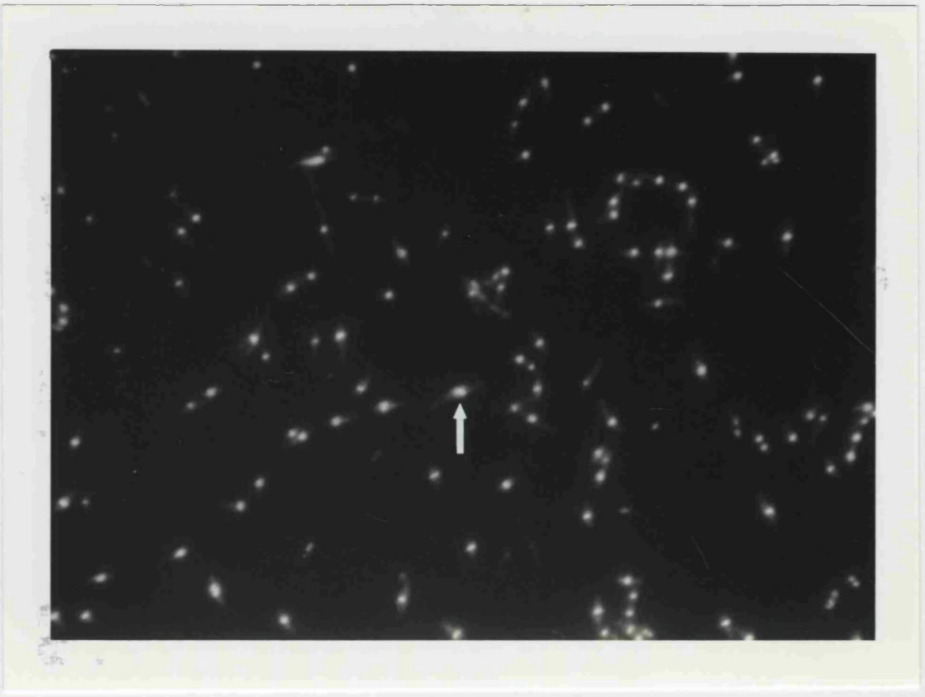
DAPI stained INVαF' and KM4104 cells expressing MBP-Ycf 24.

A. INVαF' (t = 3h).

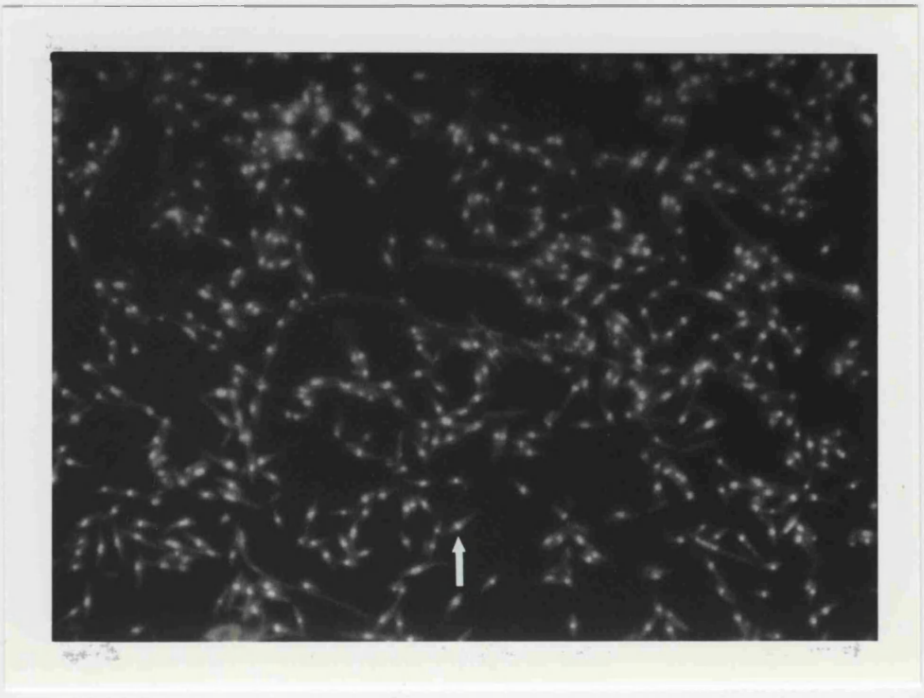
B. KM4104 (t = 3h).

Both sets of cells show moderate filamentation and condensed nucleoids highlighted by the arrows.

A



B

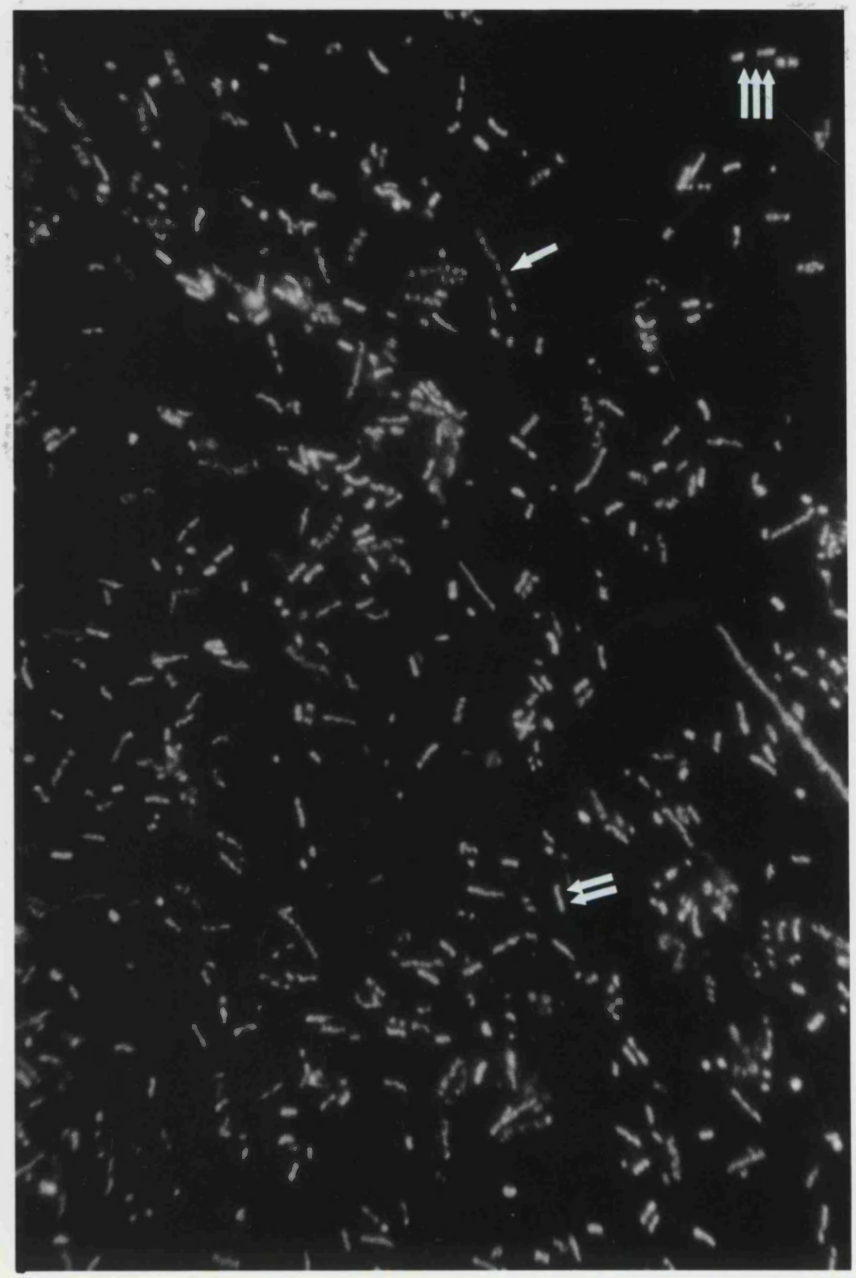


6.6μm

Figure 3.4.8

DAPI stained INVαF′ cells following three hours of induction of pMAL-c2.

The nucleoids, rather than condensed, are distributed in a row, sometimes along the entire length of the cell (single arrow), sometimes towards the centre of the cell (double arrow) and sometimes at either ends of the cell (triple arrow).

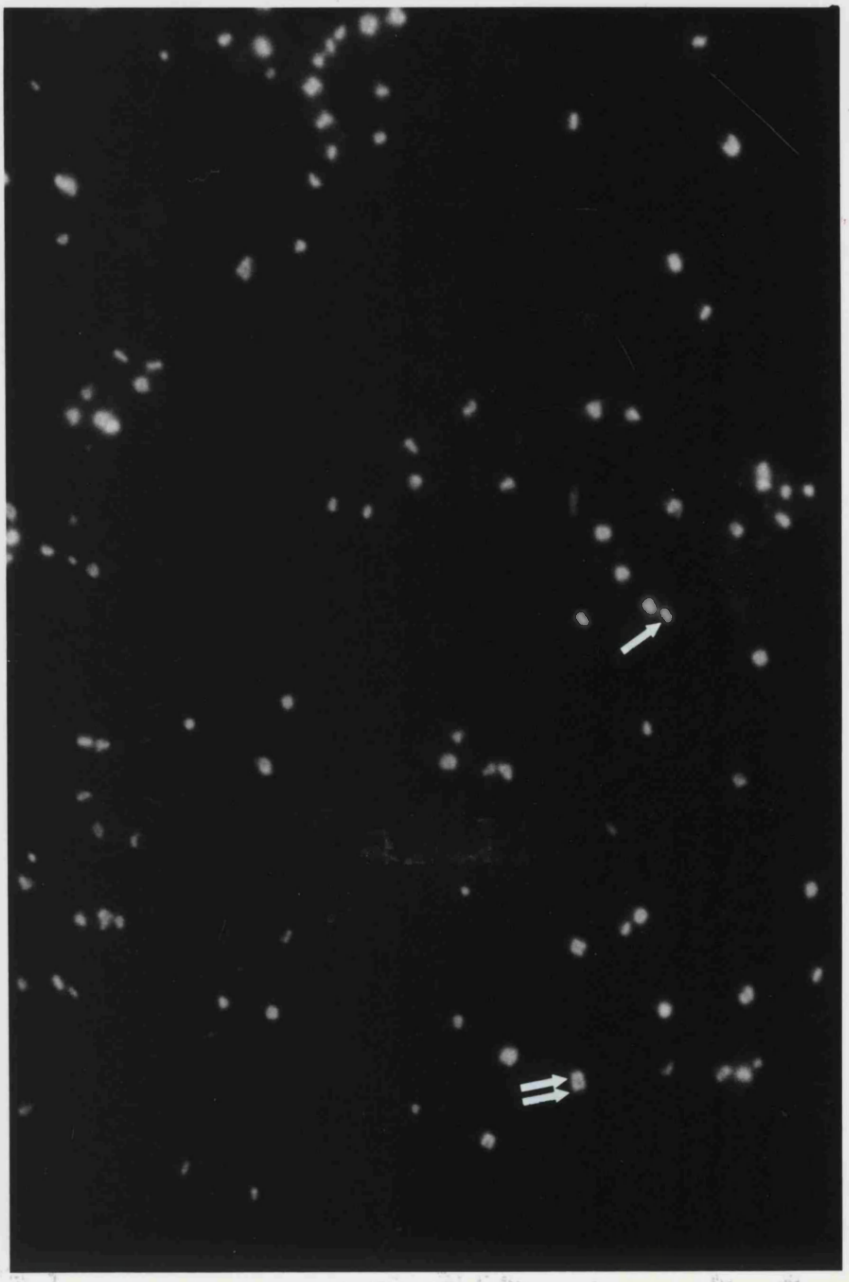


4.4μm

Figure 3.4.9

Nucleoid appearance of INVαF′ cells following three hours of induction of *ycf*24-pMAL-c2.

The single arrow highlights the dumbbell shape of the nucleoids following the first division which occurred transverse to the cell's length. The double arrow highlights the double dumbbell shape produced with a second, longitudinal division. The nucleoids however remain condensed.



4.4 μm

Figure 3.4.10

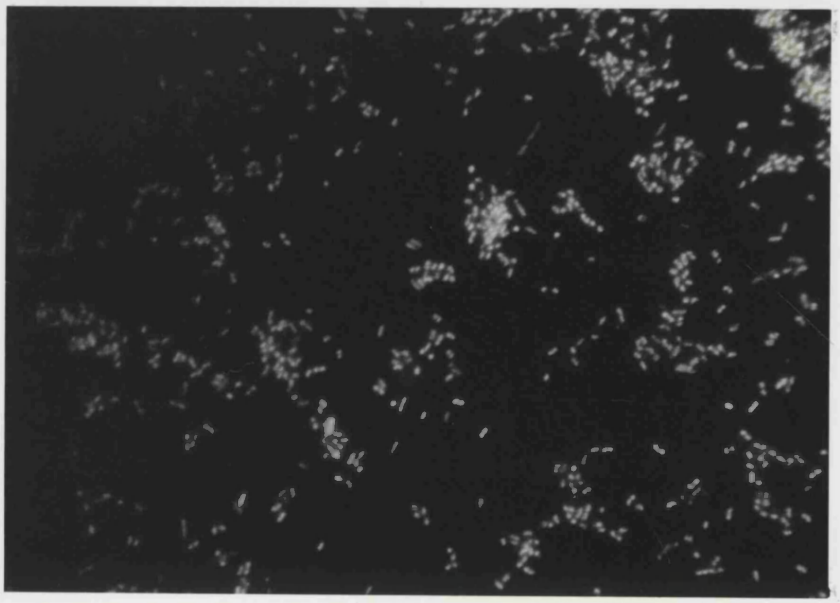
DAPI stained nucleoids of INVαF′ cells following over-expression of pMCK113 cloned into pMAL-c2.

A. pMCK113-pMAL-c2 uninduced control.

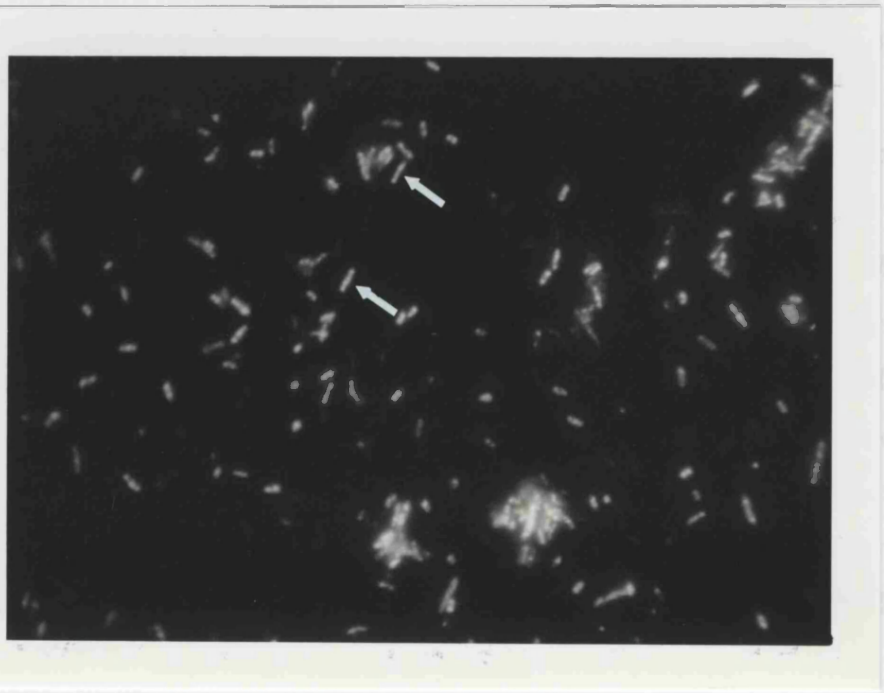
B. pMCK113-pMAL-c2 (t = 3h).

Athough cell death is apparent after induction for 3 hours, filamentation is observed in the still viable cells. The nucleoids appear normal, i.e. distributed along the cell length (arrows).

A



B



6.6μm

and A16. The latter was designed to include both a *Pst* I restriction site and a stop codon to halt expression at the end of the cloned insert.

A 16: GAG CTC **CTG CAG** TTA ATA AGT AGT GGC AAC CGA AAC TGA
(The *Pst* I site is shown in bold and the stop codon is underlined)

Following digestion with *Eco* RI and *Pst* I, truncated *ycf* 24 (T^1 *ycf* 24) was cloned into pMAL-c2 and checked for the presence of the correct sized insert by restriction digest with both *Eco* RI and *Pst* I. Clones containing the correct sized insert were sequenced and one of these (clone 10), was over-expressed as an MBP fusion protein upon induction with 0.3mM IPTG. Optical densities at A_{600} indicated expression of a fusion protein, the drop in growth rate being greater after three hours induction than with over-expression of the full length fusion protein from pMAL-c2 (Table 3.4.4). Over-expression of the truncated fusion protein (~59kDa) was confirmed by observation on Coomassie stained SDS-PAGE gels (Fig. 3.4.11, lane 8) and western blots using an antibody against MBP.

Although moderate filamentation was found to occur in cells over-expressing this truncated gene, condensed nucleoids were also apparent. In fact the nucleoids appeared more condensed compared with nucleoids of uninduced, host and vector only controls which appeared normal. An example of uninduced and induced cells containing T^1*ycf* 24-pMAL-c2 is shown in Figure 3.4.12.

A second truncation (T^2), leaving only the 258bp 5′ end of *ycf* 24 to be expressed was amplified by the primers A11 and A17. Again the reverse primer was designed to stop expression at the end of the insert (see Appendix 3). Following digestion with *Eco* RI and *Pst* I, the construct was cloned into pMAL-c2 and sequenced. Four clones were induced with 0.3mM of IPTG and samples taken t = 0, 1 and 3 hours after induction. The growth rates, which followed the same pattern as the first truncation, indicated expression of a fusion protein of approximately 51kDa. A Coomassie stained gel and western blot showing the over-expression of MBP- T^2 Ycf 24 is shown in Figure 3.4.11, lane 10.

Samples of over-expressing cells containing the second truncation were prepared and stained with DAPI and observed by fluorescence microscopy. This time, unlike the Ycf 24 and T^1Ycf 24 fusions, the nucleoids appeared like those of the cells

Optical density (A_{600})		t = 0h	t = 1h	t = 3h
INVαF'	U	0.322	0.765	1.286
	I		0.760	1.300
Vector	U	0.330	0.710	1.270
	I		0.732	1.246
MBP-Ycf 24	U	0.276	0.644	1.263
	I		0.693	1.125
MBP-T^{I}Ycf24	U	0.240	0.558	1.226
	I		0.588	1.068

Table 3.4.4

Optical densities (A_{600}) of uninduced (U) and induced (I) cultures containing recombinant pMAL-c2 plasmids or no plasmid.

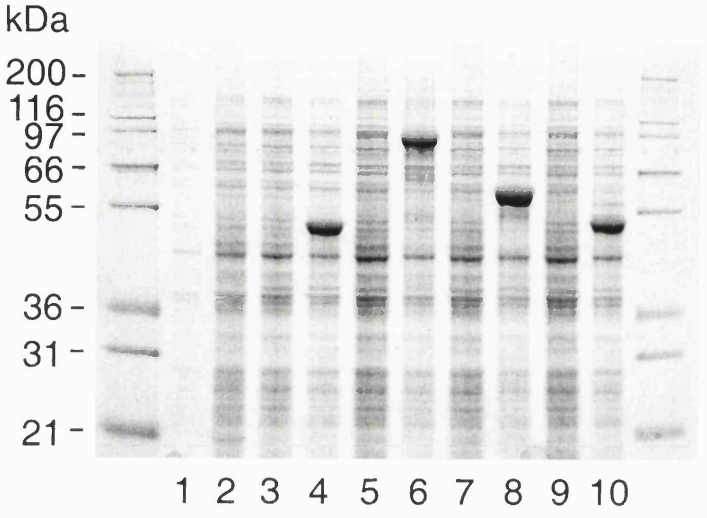


Figure 3.4.11

Western blot and Coomassie stained SDS-PAGE gels of INVαF′ cells expressing MBP fusions.

Lane: 1. t = 3hU Host control.
2. t = 3hI.
3. t = 3hU pMAL-c2.
4. t = 3hI.
5. t = 3hU *ycf* 24-pMAL-c2.
6. t = 3hI.
7. t = 3hU T^1*ycf* 24-pMAL-c2.
8. t = 3hI.
9. t = 3hU T^2*ycf* 24-pMAL-c2.
10. t = 3hI.

Lane 4 shows induction of the MBP-LacαZ (~52kDa), lane 6 MBP-Ycf 24 (~100kDa), lane 8 MBP-T^1Ycf 24 (~59kDa) and lane 10 MBP-T^2Ycf 24 (~51kDa).

Figure 3.4.12

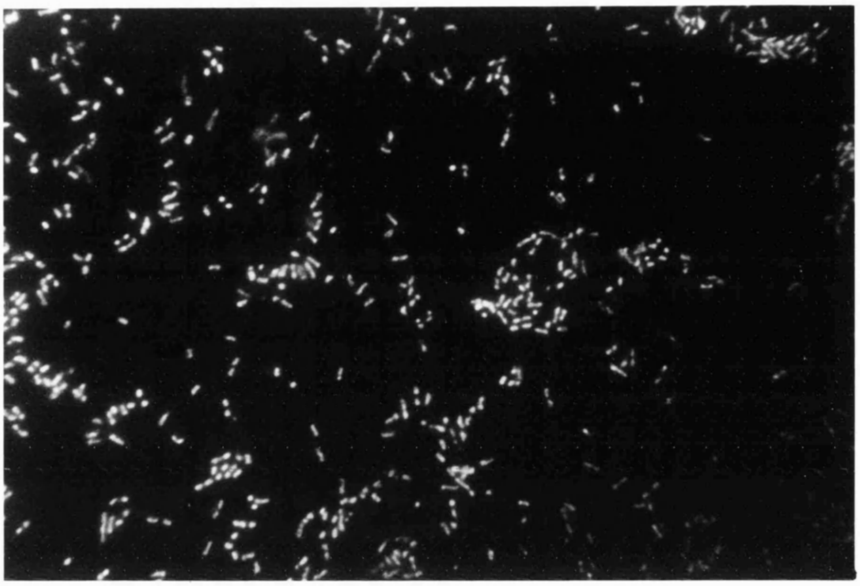
DAPI stained nucleoids of INVαF′ cells following over-expression of the first truncation of *ycf*24.

A. Uninduced control.

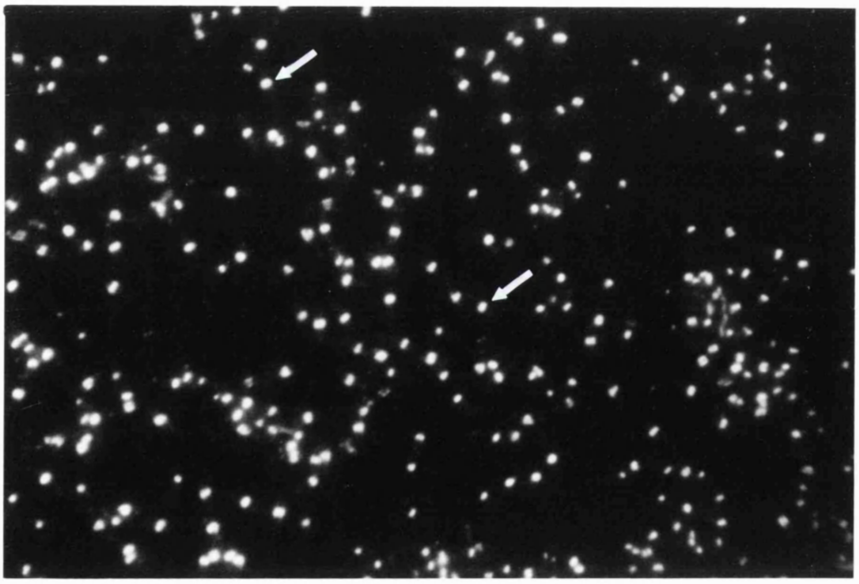
B. T[1]*ycf*24-pMAL-c2 (t = 3h).

Filamentation and condensed nucleoids are observed causing cell distortion (arrows).

A



B



6.6μm

expressing pMAL-c2 only despite filamentation. Rather than distended cells containing condense nucleoids the cells contained multiple nucleoids disributed along the cells' length (Fig. 3.4.13). These results indicated that expression of more than just the first 86 N-terminal amino acids of *ycf* 24 is required for aberrant segregation of nucleoids.

3.4.5 Localization of Ycf 24 fusion proteins

(Work done by R.J.M. (Iain) Wilson)

The *E.coli* Ycf 24 fusion protein was localized in permeabilised cells using a monoclonal antibody to the maltose binding protein (MBP) tag.

Uninduced transformants showed a small immunofluorescent polar dot (~~single arrow in Figure 3.4.14~~). Following one hour of induction the enlarged nucleoids were bounded by transverse bands of the fusion protein. The fluorescent signal was sometimes shown to extend along the membrane to each pole to delineate a 'compartment'. Longer cells showed multiple compartments with alternating nucleoids. The nature of these bands remains to be defined but they appeared to be associated with the nucleoids or fragments of DNA (DAPI staining) as opposed to occurring at irregular intervals along the cell (Fig. 3.4.14).

Over-expression of MBP-T^1*ycf* 24 (or deletion of the C′ terminal 2/3rds of *ycf* 24) gave a fusion protein that produced only a subtle change in the banded immunofluorescence pattern. It was found that the bands around the nucleoids were tighter which agreed with the increasingly aberrant partitioning of the nucleoids. Upon over-expression of the second deletion of *ycf* 24 as an MBP fusion protein (T^2) the distinctive "zebra" patterning was not observed. Induction of the maltose binding protein per se has been reported to form polar caps (Dietzel *et al.*, 1978). When pMAL-c2 alone was induced we observed an overall fluorescence inter-spaced with random, irregular and dense accumulations of MBP.

Confirmation that the "zebra" bands reflected deposition of the Ycf 24 fusion protein around the nucleoid was shown by the loss of signal after absorbing the antiserum with MBP bound to amylose resin.

Figure 3.4.13

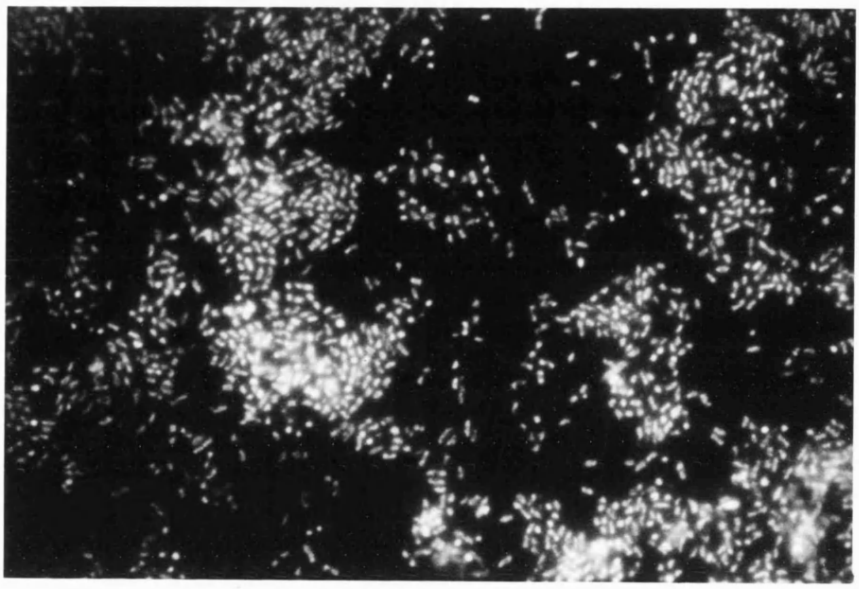
DAPI stained nucleoids of INVαF′ cells following over-expression of the second truncation of *ycf*24.

A. Uninduced control.

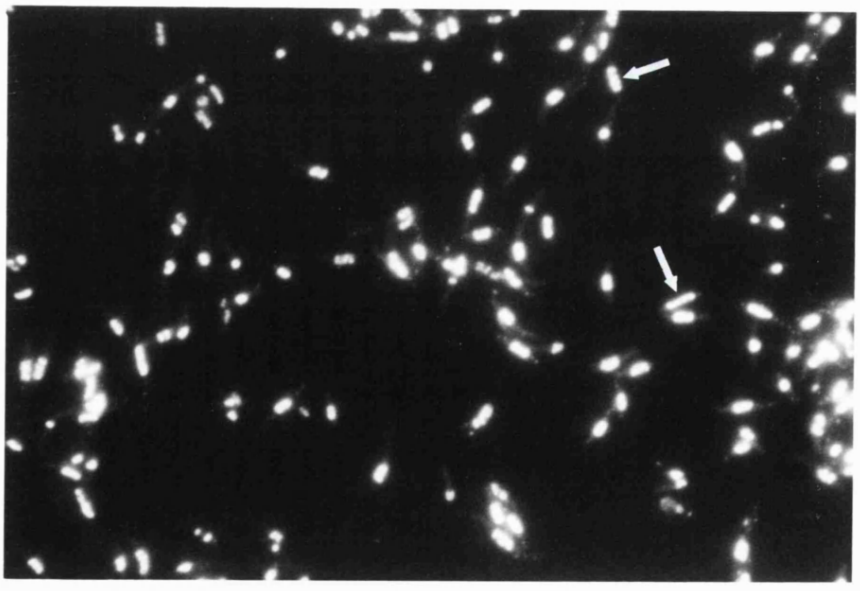
B. T^2*ycf*24-pMAL-c2 (t = 3h).

Filamentation is observed but the nucleoids appear more normal, i.e. distributed along the cells length (arrows).

A



B



5.4μm

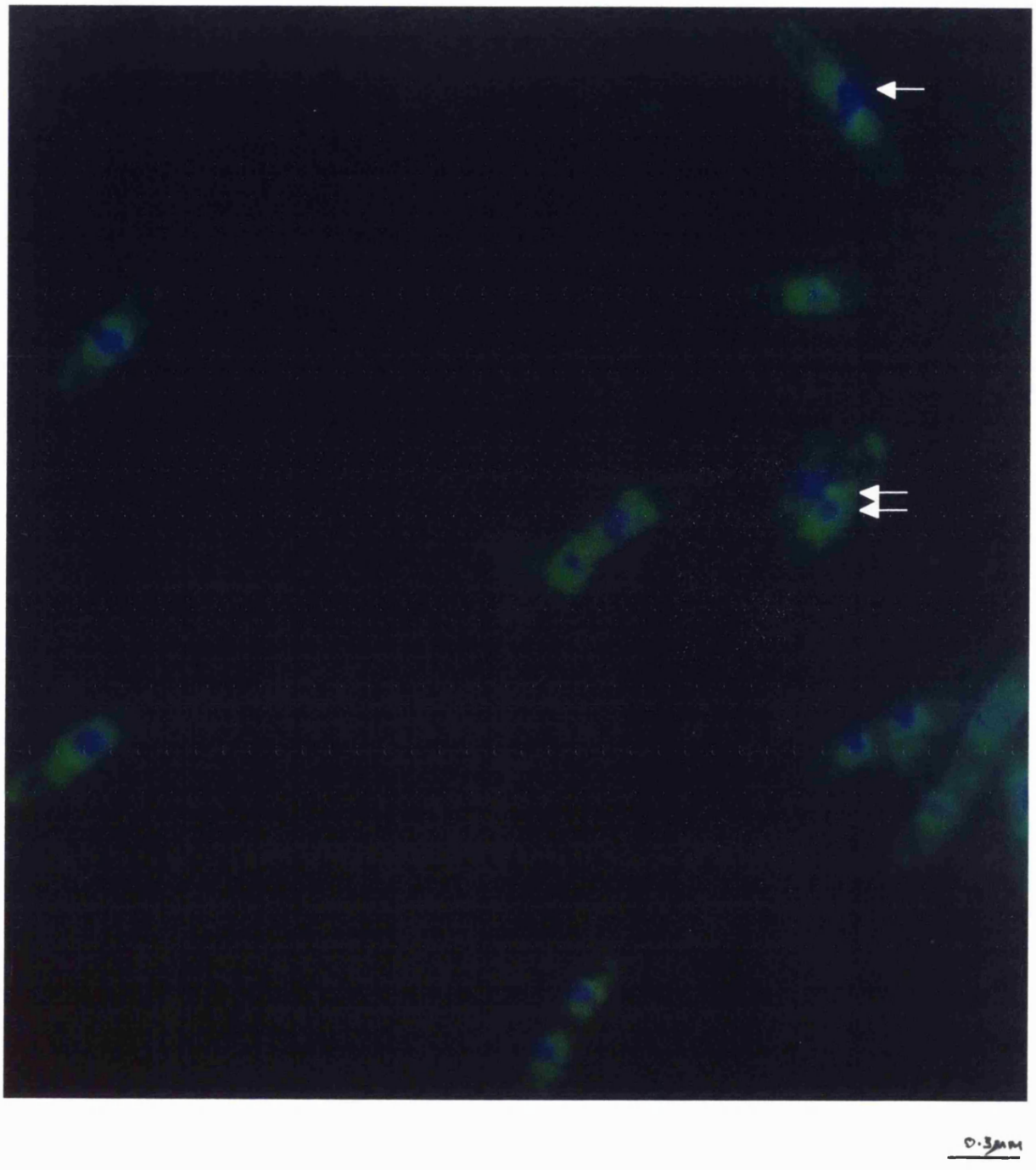


Figure 3.4.14

Localization of MBP-Ycf 24 using an antibody to MBP and detected with a fluorescent antiglobulin.

The large DAPI-stained dot (corresponding to the nucleoids) is shown in blue (single arrow).

Transverse bands of fusion protein are shown in green (double arrows). They appear to be associated with the nucleoids and the cell membrane.

3.4.6 FtsZ ring formation in Ycf 24 fusion protein expressing cells

(Fluorescent antibody localization carried out by Iain Wilson)

The "zebra" patterning of cells over-expressing Ycf 24 as a fusion protein, their filamentation, as well as the obvious effects on nucleoid segregation led me to investigate the affects on the *E. coli* cell division protein FtsZ. FtsZ is required for the contractile ring and septum formation in dividing bacteria.

Reports for other partition mutants using a fluorescent antibody against FtsZ showed multiple contractile rings (Yu and Margolin 1999). In the present study, staining of the contractile ring was only apparent in cells not expressing *ycf* 24 as a fusion protein (Fig. 3.4.15), i.e. FtsZ was prominent in host cells and uninduced cells. The contractile ring was also found to be absent from cells over-expressing T^1Ycf24. However, it was also noted that FtsZ contractile rings were not observed in filamenting cells over-expressing pMAL-c2 only. This suggested that loss of FtsZ ring formation was directly attributed to over-expression of MBP as opposed to over-expression of Ycf 24. Even so I was interested to determine if FtsZ was present in these filamenting cells.

To investigate FtsZ in these cells, western blots using an antibody against *E. coli* FtsZ were carried out on protein extracts from cultures inducing all of the Ycf 24 fusions, as well as vector only and host cell controls. All samples tested were found to contain a single 40kDa protein corresponding to FtsZ. Although a slight decrease in the amount of protein was observed in the induced samples, as judged by intensity, this was attributed to the slight reduction in growth rates for cultures expressing Ycf 24, T^1Ycf 24 and T^2Ycf 24 maltose binding protein fusions as well as for pMAL-c2 only. An example of one of these western blots (MBP-Ycf 24) is shown in Figure 3.4.16. The presence of FtsZ in cells over-expressing *ycf* 24 did not explain the loss of septum formation. Therefore, western blots were prepared using the same samples, run on native gels, and probed with the FtsZ antibody to see if the fusion proteins had bound to endogenous FtsZ causing it to run differently in native gels. However, it was found that as mobility of FtsZ was unaffected in cells expressing the Ycf 24 fusion proteins when compared to the controls (Fig. 3.4.17). It is not known if the inhibition of contractile ring formation caused by over-expression of Ycf 24 is the same as that caused by over-expression of MBP.

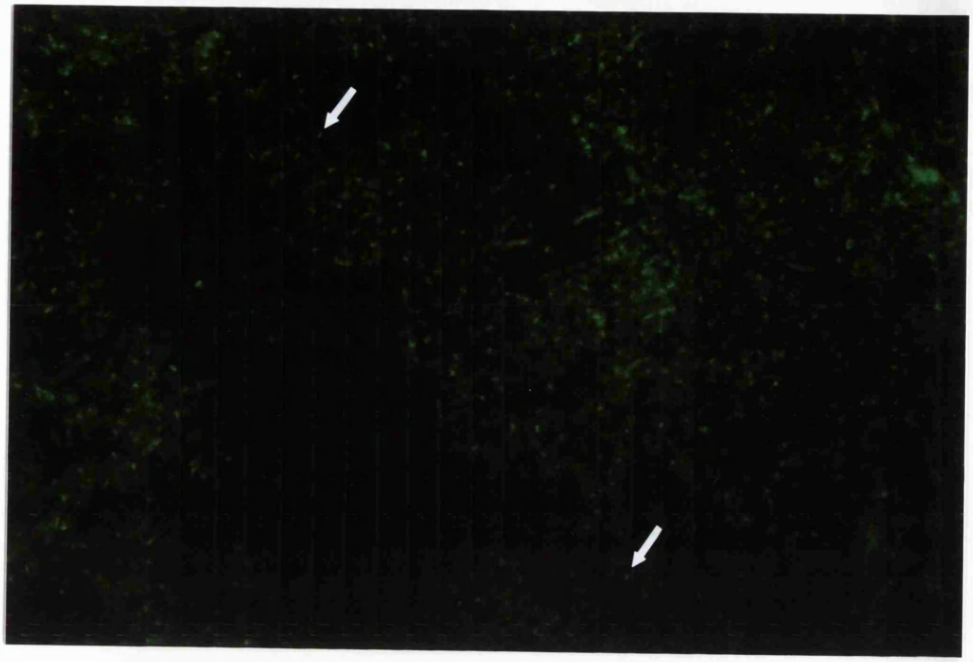
Figure 3.4.15

FtsZ contractile ring localization using an antibody to *E. coli* FtsZ.

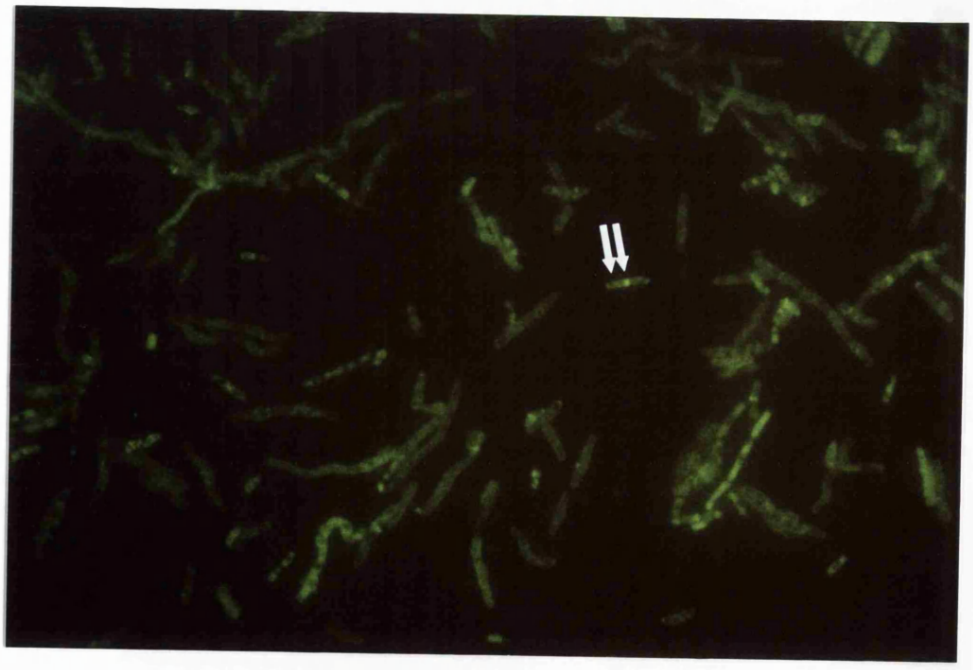
A. Uninduced control. Contractile rings are observed in the majority of the cells (single arrows).

B. *ycf* 24-pMAL-c2 (t = 3hI). Staining of the contractile ring is absent, with the exception of very few cells (double arrow).

A



B



6.6μm

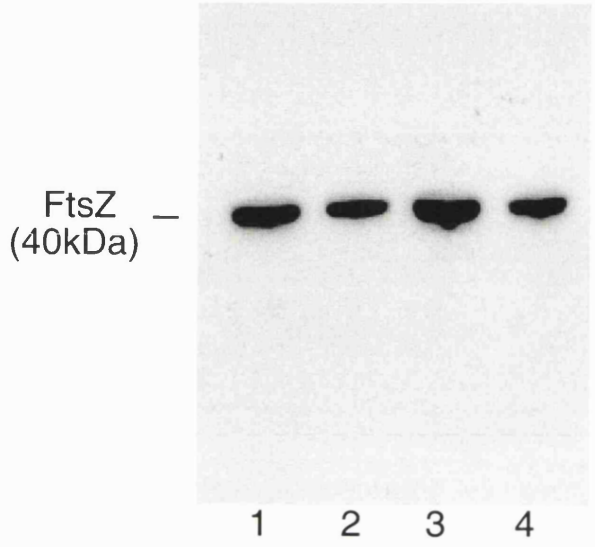


Figure 3.4.16

Western blot on total protein extracts from INVαF′ cells containing *ycf* 24-pMAL-c2 using an antibody to FtsZ.

Lane: 1. t = 1hU.
2. t = 1hI.
3. t = 3hU.
4. t = 3hI.

FtsZ (40kDa) is present in cells over-expressing MBP-Ycf 24 as well as cells not expressing the fusion protein.

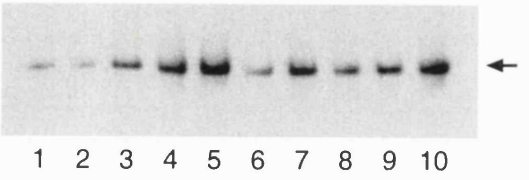


Figure 3.4.17

Native western blot on total protein extracts from INVαF′ cells using an antibody to FtsZ.

Lane: 1. t = 3hI INVαF′ host.
2. t = 3hU.
3. t = 3hI pMAL-c2.
4. t = 3hU.
5. t = 3hI *ycf* 24-pMAL-c2.
6. t = 3hU.
7. t = 3hI T^1*ycf* 24-pMAL-c2.
8. t = 3hU.
9. t = 3hI T^2*ycf* 24-pMAL-c2.
10. t = 3hU.

The mobility of FtsZ is the same in cells expressing MBP-Ycf 24 fusions as in the uninduced, host and vector only controls.

3.4.7 Conclusion

In conclusion, over-expression in *E. coli* of full length Ycf 24 as a maltose binding protein fusion resulted in moderate filamentation, loss of FtsZ ring formation and clumped nucleoids (only the last was not seen in pMAL-c2 controls). The fusion protein was found to localize as a band on either side of the condensed nucleoids. This observation indicates a possible role for *ycf* 24 in nucleoid partitioning, confirming results seen in the *Synechocystis* partial knockout.

To elucidate the function of Ycf 24 further, purification of *E. coli* Ycf 24 was attempted with a view to producing a monoclonal antibody to investigate protein/protein interactions. This is covered in the next and final results section.

3.5 Purification of *E. coli* Ycf 24

3.5.1 Purification of *ycf* 24 maltose binding protein fusions

1 litre of L-broth was inoculated with 3ml of overnight culture of *ycf* 24-pMAL-c2 in INVαF′ (clone 2, see Chapter 3.4) and grown up to an optical density (A_{600}) of approximately 0.5. IPTG was added to a final concentration of 0.3mM and induction allowed to proceed for three hours. The induced cells were centrifuged and the pellet resuspended in a smaller volume of sample buffer containing a cocktail of 1x protease inhibitors (Calbiochem®) before freezing at -20°C overnight. Proteins were released from the cells after thawing at room temperature and sonication. The soluble fraction was separated by ultra-centrifugation (50, 000 rpm for 1 hour), and a sample of (crude extract) was separated on SDS PAGE gels beside samples of uninduced and induced total cell extracts. The stained gels, and western blots probed with an antibody against MBP indicated that there was a reasonable amount of soluble fusion protein in cells over-expressing MBP-Ycf 24 (lanes 1-3 of Figures 3.5.1A, Coomassie gel, and 3.5.1B, western blot).

The soluble fusion protein was separated from other *E. coli* proteins by passage through an amylose resin column, which bound MBP and thus the Ycf 24 fusion protein. The fusion protein was then eluted into 1.5ml fractions and UV absorbance at 820nm showed it eluted as a peak between fractions 7 and 12. Samples of fractions 6-11 were run on SDS PAGE gels and stained with Coomassie blue or blotted onto nitrocellulose membrane and probed with an antibody to MBP. The results of these manipulations are shown in Figures 3.5.1A and 3.5.1B, lanes 4 to 9. They indicate the full length fusion protein was bound to and eluted from the amylose resin, although a number of contaminating proteins were also present. It was not clear how many of these were breakdown products or unrelated *E. coli* proteins. As the end result was some purification of Ycf 24, the next step of purification was attempted – cleavage of the maltose binding protein tag from Ycf 24.

3.5.2 Factor Xa cleavage of the *ycf* 24 MBP fusion protein

A small sample of the soluble crude extract isolated from cells over-expressing *ycf* 24 as a fusion protein was subjected to cleavage by factor Xa at an enzyme:protein ratio of 1:100 (w/w) (e.g. 1mg of factor Xa was used to cleave 100mg of fusion protein). Cleavage was measured over a three day incubation period at 4°C.

Figure 3.5.1

SDS-PAGE gels of MBP-Ycf 24 fusion purification.

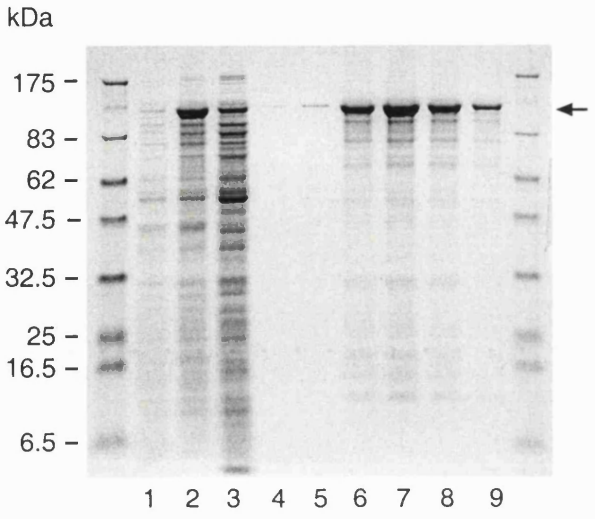
A. Coomassie stained gel.

B. Western blot using an antibody against the MBP tag.

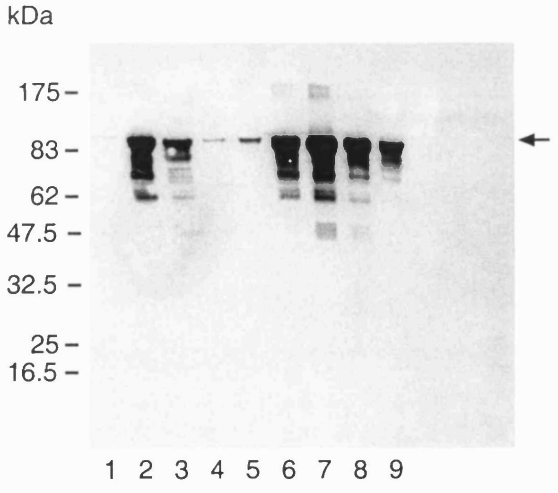
Lanes: 1. Uninduced total cell extract.
2. Total cell lysate (t = 3hI).
3. Crude extract (soluble fraction).
4. Elution fraction 6.
5. Elution fraction 7.
6. Elution fraction 8.
7. Elution fraction 9.
8. Elution fraction 10.
9. Elution fraction 11.

Partially purified MBP-Ycf 24 was found to elute from an amylose resin column between fractions 7 and 11(arrow). Also observed are additional proteins which could be breakdown products or contaminating host proteins.

A


kDa
175
83
62
47.5
32.5
25
16.5
6.5
1 2 3 4 5 6 7 8 9

B


kDa
175
83
62
47.5
32.5
25
16.5
1 2 3 4 5 6 7 8 9

No cleavage was observed so the enzyme:protein ratio was increased to 1:20 (w/w) but without success over three days at 4°C. Because of this, it was assumed the three dimensional conformation of the Ycf 24 MBP fusion prevented protease access to the cleavage site.

To overcome this, small amounts of SDS, ranging from 0.005 to 0.05%, were added to the reaction mix and incubated at 4°C overnight. This time cleavage occurred in reactions containing more than 0.010% of SDS (Fig. 3.5.2, lanes 6 to 9). At a concentration of 0.05% no full length fusion protein was observed. Two bands of approximately 60kDa and 43kDa were expected for Ycf 24 and MBP respectively. The unexpected sizes of many of the cleavage products found was assumed to be due to degradation products which had or had not been cleaved by factor Xa. However, when a western blot probed with an antibody against MBP was analysed, all the major bands were recognised (Fig. 3.5.3). This indicated that cleavage had occurred at other sites within the fusion protein.

When factor Xa cleavage was repeated using the partially purified Ycf 24 fusion protein, cleavage was not achieved even with 0.05% of SDS in the reaction. Increased concentrations of 0.075 and 0.1% of factor Xa in overnight digests still resulted in some full length fusion protein remaining. Western blots confirmed that correct cleavage was not occurring when again all the bands resulting from the cleavage were recognized by the MBP antibody. In another attempt to purify some of Ycf 24, T^{1}*ycf* 24 was over-expressed as a maltose binding protein fusion and purified by passing through amylose resin (Fig. 3.5.4). Factor Xa cleavage was again unsuccessful (Fig. 3.5.5). Before moving to another system it was decided to clone in another proteolytic cleavage site just upstream of *ycf* 24 in pMAL-c2 to see if the problem of cleaving MBP could be solved.

3.5.3 Cloning the PreScission protease site into *ycf* 24-pMAL-c2

The new protease site chosen was the PreScission protease site of pGEX-6P1 (Pharmacia Biotech.). Primers were designed which spanned the entire site and *Eco* RI sites were incorporated at either end. One of these primers 6P1-1 is shown below with the *Eco* RI sites highlighted. The second primer (6P1-2) was designed for the complimentary strand.

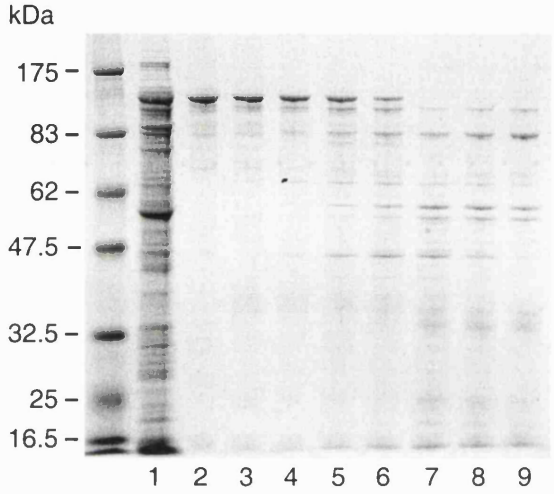


Figure 3.5.2

Coomassie stained gel following Factor Xa cleavage of MBP-Ycf 24 (SDS titration).

Lanes: 1. Crude extract (soluble fraction).
2. No factor Xa.
3. Factor Xa only.
4. 0.005% SDS.
5. 0.010% SDS.
6. 0.020% SDS.
7. 0.030% SDS.
8. 0.040% SDS.
9. 0.050% SDS.

Factor Xa cleavage is apparent in reactions containing more than 0.010% SDS.

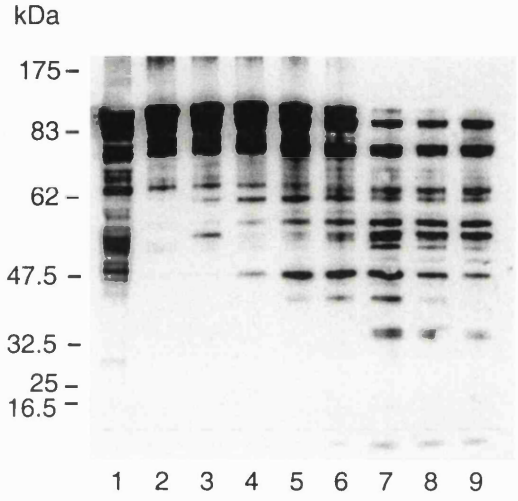


Figure 3.5.3

Western blot using anti-MBP following factor Xa cleavage of MBP-Ycf 24 (SDS titration).

Lanes: 1. Crude extract (soluble fraction).
2. No factor Xa.
3. Factor Xa only.
4. 0.005% SDS.
5. 0.010% SDS.
6. 0.020% SDS.
7. 0.030% SDS.
8. 0.040% SDS.
9. 0.050% SDS.

Although factor Xa cleavage has occurred, all products are recognized by anti-MBP indicating incorrect cleavage had occurred.

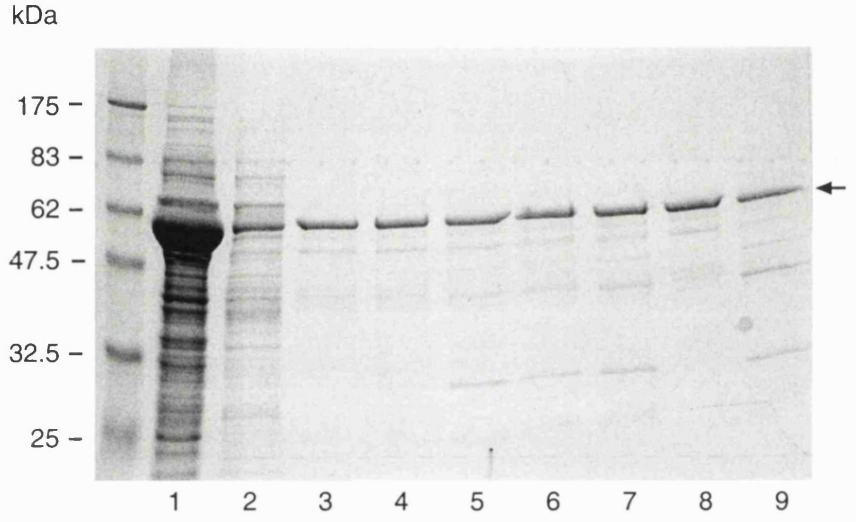


Figure 3.5.4

Coomassie stained SDS-PAGE gel following purification of MBP-T[1]Ycf 24.

Lanes: 1. Total cell lysate (t = 3hI).

2. Crude extract (soluble fraction).

3-9. Elution fractions.

Soluble MBP-T[1]Ycf 24 is apparent in all elution fractions as are additional proteins. These are either breakdown products or contaminating host proteins.

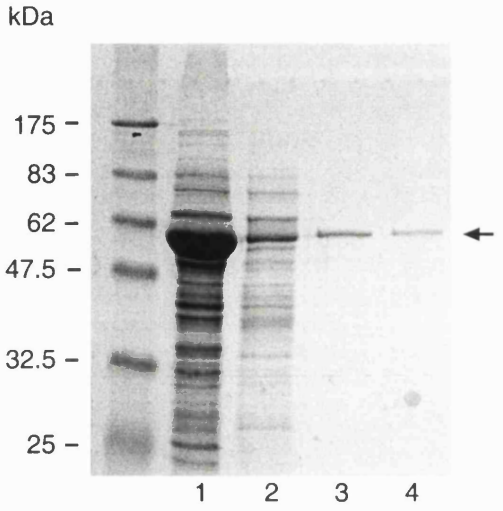


Figure 3.5.5

Coomassie stained SDS-PAGE gel following factor Xa cleavage of MBP-T[1]Ycf 24.

Lanes: 1. Total cell extract (t = 3h).
2. Crude extract (soluble fraction).
3. Factor Xa only.
4. 0.005% SDS.

No cleavage of MBP-T[1]Ycf 24 is apparent in the presence of factor Xa (lanes 3 and 4).

6P1-1

CAG CTG **GAA TTC** CTG GAA GTT CTG TTC CAG GGG CCC **GAA TTC** GAG CTC

10ng of each primer were denatured by incubation at 95°C for thirty seconds before annealing at 65°C for five minutes. After digestion with *Eco* RI the PreScission protease site was ligated into the *Eco* RI site of *ycf* 24-pMAL-c2 (the same site into which the 5′ end of *ycf* 24 was cloned). The aim was to produce recombinant clones with pMAL-c2 sequence coding for MBP, followed by the PreScission protease site and then *E. coli ycf* 24, as shown below. The pMAL-c2 sequence is shown in lain font, *Eco* RI sites are underlined, the PreScission protease site is shown in bold and the *E. coli ycf* 24 sequence is in italics. However when 10 recombinant clones were sequenced none contained the proteolytic site.

...ATC GAG GGA AGG ATT TCA GAA TTC **CTG GAA GTT CTG TTC CAG GGG CCC** GAA TTC *ATG TGG CTG TGG CGA*...

After the failure to clone an additional proteolytic site into *ycf* 24-pMAL-c2, I decided use an alternative expression vector, pET-28a.

3.5.4 Over-expression of *ycf* 24 as a histidine tagged fusion protein

Over-expression of genes cloned into pET-28a allows the production of a fusion protein with histidine (His) tags at both the N and C-termini. The N-terminal tag can subsequently be removed by thrombin cleavage. As only one tag is required to allow purification by specific binding and elution from either a zinc or nickel column, it was decided to clone *ycf* 24 into the *Nco* I site within pET-28a before the N-terminal histidine and thrombin cleavage coding region. This would result in a fusion protein carrying a small C-terminal hexa-His tag which might not interfere with the conformation of the protein. It also avoided the cleavage of the N-terminal His tag for purification of the protein following over-expression. Two primers were designed to do this. The forward primer (pET-28a1, shown below) allowed the cloning of *ycf* 24 into the *Nco* I site (underlined) using ATG as the start codon (in bold), not the *ycf* 24 ATG. An extra two bases (G and A, shown in brackets) were incorporated in the

primer to ensure the sequence was in frame- resulting in an additional glycine at the N-terminus (italics). The reverse primer, pET-28a2 (also shown below), was designed to clone into the *Xho* I site (bold) in frame with the 3′ histidine coding sequence, it therefore contained no *ycf*24 stop codon.

pET-28a1
GAG CT CCAT G*G(G A)*TG GCT GTG GCG AAA G

pET-28a2
GAG CTC **CTC GAG** TCC GAC GCT GTG TTC A

Following amplification using these primers from SURE DNA, *ycf* 24 was digested with both *Nco* I and *Xho* I and cloned into pET-28a. Recombinants in BL21(DE3)pLysS cells were selected by growth on kanamycin. Only twelve colonies were observed. However, nine were found to contain a correct sized insert for *ycf* 24. Six of these were sequenced using various primers designed to both vector and insert sequences. One of these clones (clone 1) was over-expressed using 0.4mM IPTG. Samples were taken at t = 0, 1, 2 and 3 hours after induction. A reduction in the optical densities at A_{600} for the induced samples, when compared to the uninduced samples indicated that over-expression of the fusion protein was likely (Table 3.5.1). Expression of a fusion protein of approximately 55-60kDa was confirmed, initially by Commassie blue stained SDS PAGE gels (Fig. 3.5.6) and later by western blots using an antibody to penta-His (Fig. 3.5.7).

3.5.5 Purification under native conditions

To determine if the Ycf 24 his-tagged fusion protein was soluble it was purified under native conditions by using Ni-NTA affinity chromatography. Following lysis of a small sample (1ml) of induced cultures, one and three hours after induction, the insoluble fraction (pellet) was separated from the soluble fraction (supernatant) by centrifugation (Fig. 3.5.8). No soluble protein was detected and as a consequence none bound to the Ni-NTA agarose. This indicated the fusion protein was insoluble and needed to be purified under denaturing conditions.

Optical density (A_{600})		t = 0h	t = 1h	t = 2h	t = 3h
BL21(DE3)PlysS	U	0.497	1.191	1.549	1.718
	I		1.164	1.525	1.169
Clone 1	U	0.596	1.306	1.706	1.848
	I		1.265	1.465	1.152

Table 3.5.1

Optical density readings following over-expression of *ycf*24 cloned into pET-28a and BL21(DE3)plysS host control.

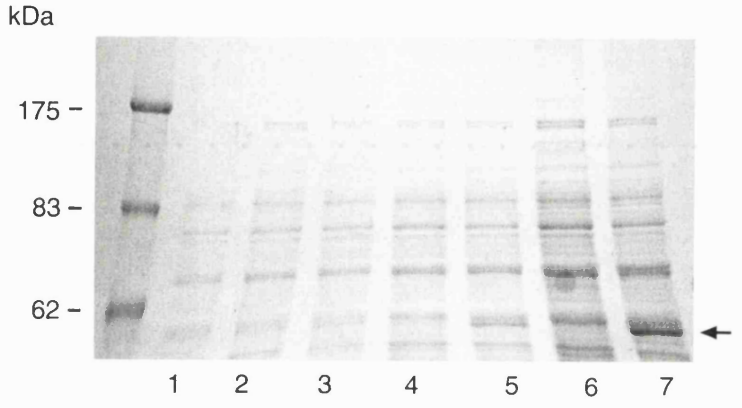


Figure 3.5.6

Coomassie stained gel following over-expression of *ycf* 24 in pET-28a.

Lanes: 1. t = 0h.
2. t = 30mins.U.
3. t = 30mins.I.
4. t = 1hU.
5. t = 1hI.
6. t = 2hU.
7. t = 2hI.

Expression of a His-tagged fusion protein of approximately 55-60kDa is shown in lanes 3, 5 and 8 (arrow).

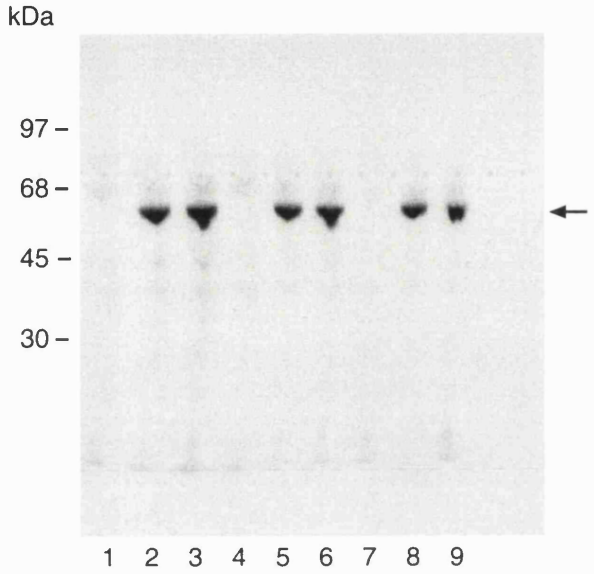


Figure 3.5.7

Western blot following induction of *ycf* 24-pET-28a using a penta-His antibody.

Lanes: 1. t = 1hU, clone 1.
2. t = 1hI.
3. t = 3hI.
4. t = 1hU, clone 2.
5. t = 1hI.
6. t = 3hI.
7. t = 1hU, clone 3.
8. t = 1hI.
9. t = 3hI.

Expression of a His-tagged fusion protein of approximately 55-60kDa is shown in lanes 2 and 3 for clone 1, 5 and 6 for clone 2, and 8 and 9 for clone 3.

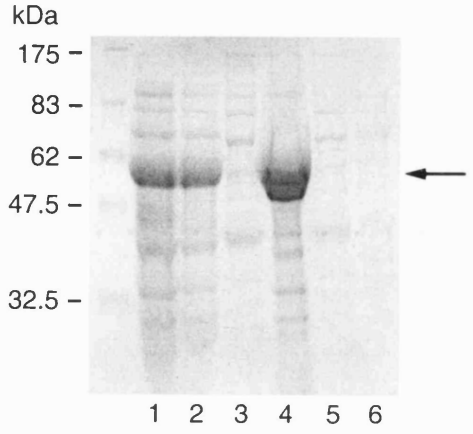


Figure 3.5.8

Coomassie stained SDS-PAGE gel following native purification of His-tagged Ycf 24.

Lanes: 1. Total cell lysate (uninduced).
2. Total cell lysate (t = 3hI).
3. Lysate supernatant (soluble protein).
4. Lysate pellet (insoluble protein).
5. Unbound supernatant.
6. Bound pellet.

This showed that the fusion protein was insoluble (lane 4) and therefore none was present in the supernatant to bind to the Ni-NTA agarose (lanes 5 and 6).

To confirm this, protein purification under native conditions was repeated using the Xpress system, a protein purification kit from Invitrogen. Here 50ml of induced culture was used but only a very small amount of soluble fusion protein was recovered as indicated in lane 3 of Figures 3.5.9A and 3.5.9B. This was applied to the nickel (ProBondTM) column and washing and elution steps carried out by the application of increasing concentrations of imidazole (from 50mM to 250mM). No fusion protein was observed in any of the fractions collected, either on Coomassie stained gels or on western blots using an antibody to penta-His (Figs. 3.5.9A and 3.5.9B, lanes 4-10).

3.5.6 Purification under denaturing conditions

Purification under denaturing conditions was carried out using the Xpress system from Invitrogen. This time, cells over-expressing Ycf 24 were lysed in the presence of guanidinium-HCl before centrifugation and separation of the solubilized proteins from the insoluble pellet. This crude extract was bound to the nickel ProBondTM column and washed with buffers containing 6M urea at pH7.8 and pH6.0. Finally the fusion protein was eluted with another urea containing buffer at pH 4.0. As expected no fusion protein was removed by either of the washes. However, small amounts of fusion protein were present in eluates 3 to 5 (lanes 4-6 in Figures 3.5.10A and 3.5.10B). This indicated that the vast majority of the *ycf* 24 fusion protein had not solubilised. Also apparent was a lot of contaminating proteins which indicated that the nickel resin was not washed sufficiently.

Since the fusion protein produced by over-expression of *ycf* 24-pET-28a was insoluble, cloning of *E. coli ycf* 24 into a third expression vector pGEX-6P1 was attempted. Although this meant that Ycf 24 would eventually have to be cleaved from its fusion protein (GST, glutathione S-transferase) it was hoped that the protein produced could be soluble and easier to cleave. It is important to note here that cells over-expressing *ycf* 24 as a His-tagged fusion protein only showed moderate filamentation. Unlike cells over-expressing *ycf* 24 as a maltose binding protein fusion the nucleoids appeared normal rather than condensed. The failure of the nucleoids to condense is attributed to the fusion protein being insoluble, unlike the Ycf 24-MBP fusion.

Figure 3.5.9

Purification of His-tagged Ycf 24 under native conditions.

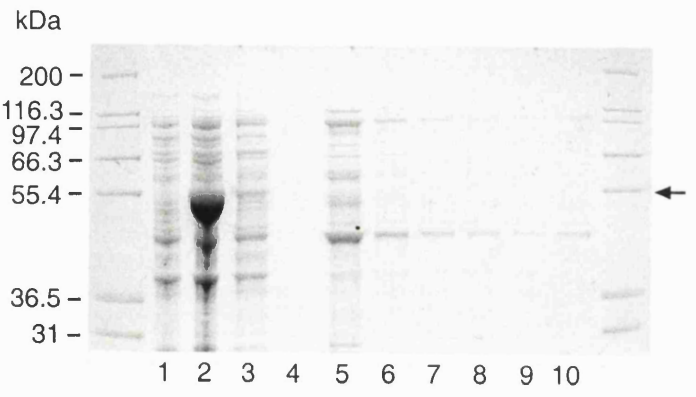
A. Coomassie stained SDS-PAGE gel.

B. Western blot using a penta-His antibody.

Lanes: 1. Uninduced total cell extract.
2. Total cell lysate (t = 1hI).
3. Crude extract (soluble fraction).
4-10. Elution fractions.

Little soluble protein was obseved (lane 3) nor was any fusion protein (arrow) seen in the elution fractions.

A


kDa
200
116.3
97.4
66.3
55.4
36.5
31
1 2 3 4 5 6 7 8 9 10

B

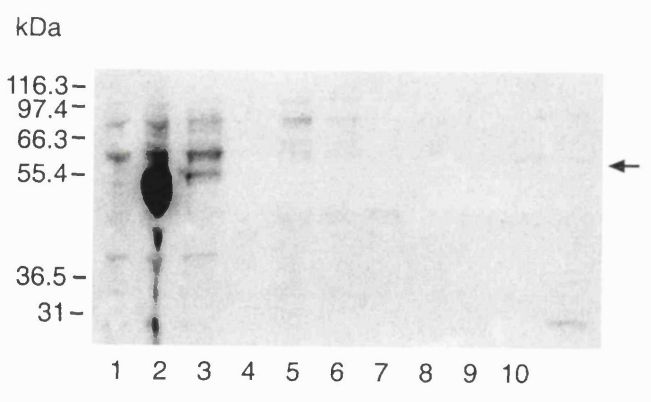

kDa
116.3
97.4
66.3
55.4
36.5
31
1 2 3 4 5 6 7 8 9 10

Figure 3.5.10

Purification of His-tagged Ycf 24 under denaturing conditions.

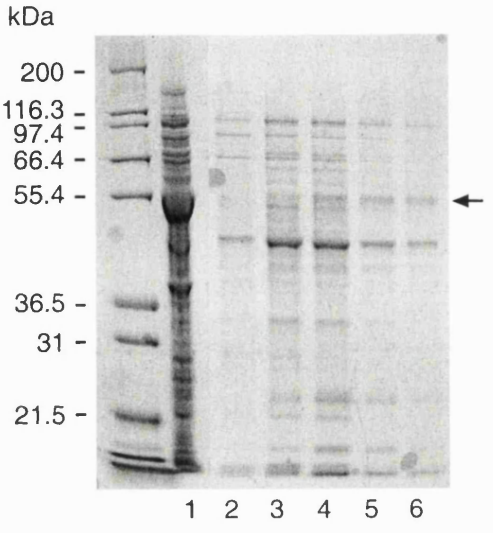
A. Coomassie stained SDS-PAGE gel.

B. Western blot using a penta-His antibody.

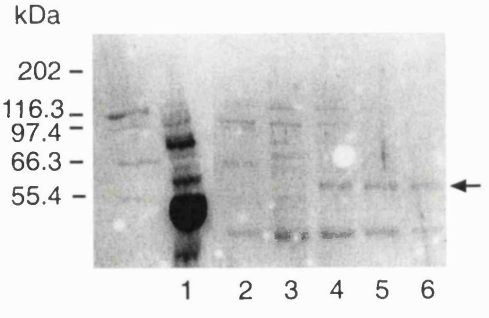
Lanes: 1. Total cell lysate (t =1hI).
2. Elution fraction 1.
3. Elution fraction 2.
4. Elution fraction 3.
5. Elution fraction 4.
6. Elution fraction 5.

Little fusion protein was found to be soluble however fusion protein was eluted in fractions 3-5 (lanes 4-6).

A


kDa
200
116.3
97.4
66.4
55.4
36.5
31
21.5
1 2 3 4 5 6

B


kDa
202
116.3
97.4
66.3
55.4
1 2 3 4 5 6

3.5.7 Over-expression of *ycf*24 as a glutathione S-transferase fusion protein

*E. coli ycf*24 was amplified from *E. coli* SURE DNA using the primers A18 and A19 which were designed to incorporate *Bam* HI and *Eco* RI sites respectively, for cloning into pGEX-6P1. Following ligation of the digested PCR product into pGEX-6P1, recombinant clones in INVαF′ cells were selected by growth on ampicillin. Ten clones were analysed by digestion with *Eco* RI and *Bam* HI and found to contain inserts of the correct size for *ycf*24 (1.5kb).

Three of these clones were induced with 0.3mM IPTG and OD readings at A_{600} were taken at t = 0, 1, 2 and 3 hours after induction. As before, a slight decrease in the optical densities indicated expression of a fusion protein which was later confirmed by Coomassie stained SDS-PAGE gels and western blots using an antibody against GST. Figures 3.5.11A and 3.5.11B show protein production in both uninduced and induced cultures after one hour. A fusion protein of the expected size (just over 80kDa) was observed, GST being 26kDa and Ycf 24 running between 55 and 60kDa.

3.5.8 Purification and cleavage of the GST fusion protein

20ml of induced culture was centrifuged and the cells re-suspended in 2ml of PreScission protease buffer before lysing and incubating with lysozyme. After sonication, the lysed cells were subjected to ultracentrifugation (50,000g for one hour) to pellet insoluble protein. A sample of the crude supernatant was incubated overnight at 4°C with 2.5U of PreScission protease. When samples of both the soluble and insoluble fractions were run on SDS-PAGE gels alongside a sample of the crude extract treated with the protease it was immediately obvious that the GST fusion protein was insoluble. Western blots using an antibody against GST were also carried out as they are much more sensitive. Again no soluble fusion protein or cleavage was apparent. The band around 46kDa corresponds to the PreScission protease (Fig. 3.5.12, lane 6).

This system resulted in a fusion protein that was both insoluble and would require cleavage of the GST tag with a protease following solubilization under denaturing conditions. Because of this it was decided to back track and to continue with the purification of the His-tagged protein (produced upon the over-expression of *ycf*24 from pET-28a).

Figure 3.5.11

SDS-PAGE analysis following induction of *ycf*24-pGEX-6P1.

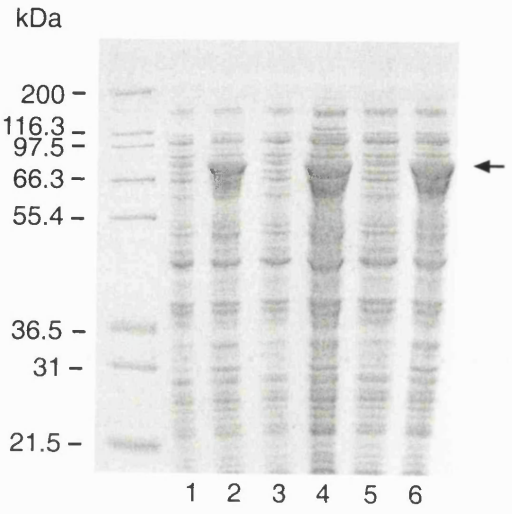
A. Coomassie stained gel.

B. Western blot using a penta-His antibody.

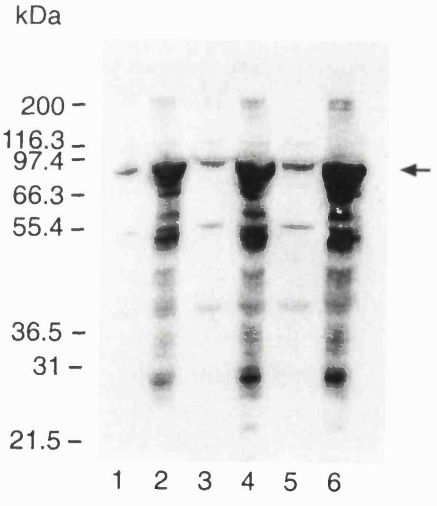
Lanes: 1. t = 0h (clone 1).
2. t = 2hI.
3. t = 0h (clone 2).
4. t = 2I.
5. t = 0h (clone 3).
6. t = 2hI.

Induction of a GST fusion protein of approximately 85kDa is observed in all of the induced samples (lanes 2, 4 and 6).

A


kDa
200
116.3
97.5
66.3
55.4
36.5
31
21.5
1 2 3 4 5 6

B


kDa
200
116.3
97.4
66.3
55.4
36.5
31
21.5
1 2 3 4 5 6

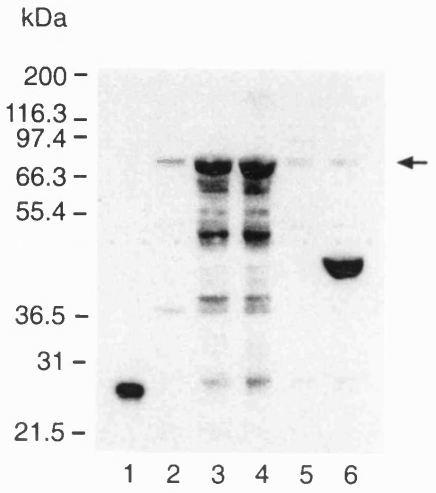


Figure 3.5.12

Western blot following a test digestion of GST-Ycf 24 with PreScission protease.

Lanes: 1. GST control (26kDa).
2. t = 0.
3. t = 2hI (clone 1).
4. Insoluble protein.
5. Concentrated crude extract.
6. PreScission protease digest.

No soluble protein was observed. Consequently no cleavage with PreScission protease occurred.

This work has been taken up by Dr Kaveri Rangachari (NIMR). So far she has successfully purified, under denaturing conditions, the fusion protein as inclusion bodies isolated from cells over-expressing *ycf* 24-pET-28a. This method described by Zhao *et al.* (1999), involved both high- and low-speed centrifugation to purify inclusion bodies. In this case, solubilization of the protein (penicillin-binding protein 2a of *Streptococcus pneumoniae*) was then achieved via a single-step refolding method. This consisted of solubilizing the inclusion bodies in urea and direct dialysis of the solubilized preparations. However, when this was attempted with purified His-tagged Ycf 24 inclusion bodies no soluble protein was obtained.

3.5.9 Conclusion

Over-expression of *ycf* 24 in *E. coli* has been achieved, as has purification of a soluble maltose binding protein fusion. However purification of Ycf 24 has been hindered by insolubility and the failure to cleave it from the maltose binding protein fusion. Nevertheless, mice have been immunised with purified MBP-Ycf 24 by Malcolm Strath (NIMR) and it is hoped that monoclonal antibodies specific to Ycf 24 may still be obtained.

Chapter 4. Discussion

4.1 Knockout in *Synechocystis* PCC6803

The failure to produce homoplasmic knockouts after long term selection led to the conclusion that *ycf* 24 is an essential gene in *Synechocystis* PCC6803. Growth of the heteroplasmic genotype was severely impaired, suggesting that the gene product is required in stochiometric amounts. The smooth mutants, which I concluded carry a compensatory mutation, were able to support a higher ratio of kanamycin-disrupted *ycf* 24 than the ragged mutants; however, homoplasmic smooth transformants were not observed. Although smooth suppressor mutants grew as well, or even slightly more vigorously than WT, they still had electron translucent plaques characteristic of the ragged mutants and a defect in septation. The presence of the plaques in a small percentage of WT cells led to the suggestion that the phenotype was characteristic of sick cells as opposed to a primary effect of reducing the number of copies of WT *ycf* 24. However, it is possible this is not the case. It was hypothesized that these plaques were either glycogen deposits or abnormal phycobilisome morphology subsequent to phycobilisome protein degradation. Nutrient deficient cyanobacteria are known to show alterations in the abundance and types of intracellular inclusion bodies, many of which serve as nutrient reserves (Allen, 1984; Simon, 1987). Glycogen, which is a storage form of fixed carbon, is visible in thin sections of cyanobacteria as electron-transparent dots. These are often located between the thylakoid membranes (Stanier, 1988) and appear very similar to those observed in the *ycf* 24 mutants. In exponentially growing cells, glycogen deposits are rarely visible whereas the cellular content of glycogen increases in iron, sulphur and phosphate deficient bacteria. It can also accumulate in nitrogen deficient cells, but not in carbon deficient cells (Grossman *et al.*, 1994). Similarly, the major light harvesting complexes can also serve as nutrient reserves, particularly in cells deprived of nitrogen, but also in times of reduced sulphur availability (Grossman *et al.*, 1993). Although Boussiba and Richmond (1980) and Wyman *et al.* (1985) have shown that phycobilisome proteins can be degraded during nitrogen stress without affecting growth rate, this is not usually the case. In general, intracellular reserves for a specific nutrient decline when that nutrient is limiting growth and as a consequence both the types and number of intracellular reserves can provide information on the nutrient status of the cell (Grossman *et al.*, 1993). Thus, the increase in the number of plaques we observed

may be an indication that cells lacking a full complement of WT *ycf* 24 are unable to assimilate some nutrient from their environment.

Phycobilisome degradation is accompanied by a massive accumulation of glycogen (Sherman and Sherman, 1983; Stevens *et al.*, 1981). Furthermore, the initial site of granule accumulation in cells undergoing stress is at the outer surface of the thylakoid membrane, a region vacated by degraded phycobilisomes (Reithman *et al.*, 1988). Consequently, it is possible that the plaques we observed are a result of phycobilisome degradation and subsequent glycogen accumulation. The levels of mRNA encoding the phycobiliproteins decline during nutrient limiting growth (Grossman *et al.*, 1993), and de Lorimer *et al.* (1984) reported that the level of phycocyanin decreased to nearly zero in nitrogen starved *Synechococcus* sp. strain PCC7002. Similarly, nitrogen-starved cultures of *Anabaena variablis* exhibit a loss of phycobiliprotein subunit mRNAs (Wealand *et al.*, 1989) as does *Synechococcus* PCC7942 under nitrogen and sulphur stress (Collier and Grossman, 1992). Therefore, it may be possible to determine if the *ycf* 24 mutants are unable to assimilate nutrients from the environment by measuring the level of phycobiliprotein gene expression in WT and *ycf* 24 mutants of *Synechocystis*. Conversely, confirmation of nutrient deprivation and loss of subunits of the light harvesting complex of *Synechocystis* might be determined using immuno-EM with antibodies specific for phycobilisomes.

As mentioned above, phycobilisome degradation and glycogen deposition is particularly prevalent under nitrogen stress. Nitrogen is required for the biosynthesis of essential amino acids involved in protein synthesis. In my experiments *Synechocystis* PCC6803 was grown routinely on BG11 plates containing nitrate as a nitrogen source. In the first step of converting nitrate to amino acids, nitrate is reduced to nitrite. I found that switching *ycf* 24 mutants from a nitrate to a nitrite source had no effect on cell growth. However, more analysis on different nitrogen sources could be carried out; alternative sources include ammonia (the second reduction step of nitrate conversion) and the amino acid L-arginine. Care must be taken when using amino acids as a source of nitrogen as many, e.g. L-glutamine, L-histidine and L-lysine cannot be used as they inhibit the growth of *Synechocystis* PCC6803 (Flores and Muro-Pastor, 1990).

Another explanation for the plaques might be the inability of cells depleted in *ycf* 24 to synthesise essential amino acids. Under these conditions phycobilisome degradation would be due to the requirement for aspartate and aromatic amino acids.

Plastids in plants play an important role in the biosynthesis of these two groups of amino acids (Lea *et al.*, 1982). As *ycf* 24 is a plastid encoded gene in *P. falciparum*, such a role could explain its maintenance on the vestigial plastid genome of apicomplexans. However, although amino acid biosynthesis in the apicoplast cannot be discounted, it seems likely that cytosolic pathways fulfil this role (McFadden *et al.*, 1997; Wilson and Williamson, 1997). Although I found supplementation of the medium with examples of both aromatic and aspartate amino acids did not improve the growth of ragged *ycf* 24 mutants, this may have been due to their inhibitory properties, as described by Flores and Muro-Pastor (1990).

Aberrant cytokinesis was observed in both the smooth and ragged mutants of *ycf* 24 following analysis by scanning electron microscopy. The specific nature of this defect was confirmed by growing the ragged mutants without antibiotic selection, which caused the number of successful cell divisions to revert to WT. Furthermore, staining with DAPI showed the nucleoids of both WT and smooth mutants contained more DNA than cells of the ragged phenotype. Our suggestion that an altered rate of cell division in cells of the ragged phenotype might be responsible for the increased number of dividing cells and reduced DNA content, was discounted after growth rate analysis of individual cells showed both the WT and *ycf* 24 mutants had a doubling time of approximately 8 hours. However, fewer cells appeared to arrest during cytokinesis in the smooth mutants than in the ragged ones, and a higher proportion of cells developed into colonies. Variation between the SEM and cell cycle data obtained for the cells of the smooth phenotype was attributed to the segregation of both the suppressor mutation and disrupted *ycf* 24. Because failed cytokinesis was not apparent in the overall appearance of the colonies of the smooth mutants, these colonies might have developed from cells carrying enough copies of the suppressor mutation to promote "normal" cell division, or to prevent cell death. More analysis of the growth rate needs to be carried out to confirm these initial results as only small numbers of cells were examined. Also, it is not known why there is less DNA in the ragged *ycf* 24 mutants hence more experiments need to be done to observe DNA segregation in these cells throughout their cell cycle; it may be that cell death during cytokinesis resulted from insufficient DNA replication or defective DNA segregation. Perhaps one of the first steps would be to measure the DNA content within individual WT and *ycf* 24 mutants of *Synechocystis* using image cytometry. This method has been used to

sucessfully measure the DNA content of individual nucleoids of *E. coli* cell division mutants (Huls *et al.*, 1999).

The mixed phenotypes discussed above make it difficult to determine the role of Ycf 24 in *Synechocystis*. Although the plaques, indicative of nutrient-deprived cells, could indicate a defect in nutrient assimilation, the predicted Ycf 24 peptide shows little homology to any proteins involved in transport (J. Saldanha, NIMR, personal communication). Furthermore, the smooth suppressor mutants still contained plaques and remained viable, failure of cytokinesis not being observed in the colony morphology (although it was present at the cellular level). If defective uptake of nutrients from the environment was responsible for cell death it seems more likely that the secondary mutation would result in a reversion of the plaque phenotype; instead it seems to have prevented cell death and aberrant cytokinesis during colony formation. I conclude that the appearance of small chains of incompletely separated cells in the smooth mutants, coupled with the results observed for over-expression of *E. coli ycf* 24, suggests this gene should be added to the list of genes affecting cell division and/or nucleoid segregation already recognised in the genome of *Synechocystis* PCC6803. These include *ftsY* and *ftsZ*; *maf*; *minC*, *minD* and *minE*; *sulA*; as well as *slr*0374 (see http//:www.kazusa, or http//:www.jp/cyano.html).

Besides the additional experiments already mentioned there are many others that need to be done to confirm some of the results observed so far. The first of these is the rescue experiment, i.e. substituting disrupted *ycf* 24 with the WT gene and selecting with a different antibiotic resistance gene. Although this experiment was attempted, retrospective analysis suggests it failed for technical reasons, i.e. absence of homologous sequence either side of the antibiotic resistance cassette; Labarre *et al.* (1989) showed that gene replacement only occurs in *Synechocystis* PCC6803 when an antibiotic resistance cassette is flanked by homologous sequence. I would also have liked to try to obtain a complete knockout in *Synechocystis* by complementing the disrupted gene with *ycf* 24 carried on the conditional expression vector, pFC1. This vector was constructed from the stable, autonomously-replicating, promiscuous plasmid RSF1010 of both *Synechocystis* and *Synechococcus* species (Mermet-Bouvier *et al.*, 1993). Included in pFC1 is the λ*c*I875 repressor-encoding gene and P_R promoter, followed by the λ*cro* ribosome-binding site and the ATG translation initiation codon. The ATG codon is located within a unique *Nde* I restriction site and

allows in-frame fusion of protein-coding sequences. Other features include genes encoding resistance to streptomycin and chloramphenicol; and the genes of RSF1010 for mobilization, origins of conjugative transfer and vegetative replication (Mermet-Bouvier and Chauvat, 1994).

In one experiment using the pFC1 vector, the *E. coli* reporter gene *lacZ* was cloned in frame and expression of β-Galactosidase was confirmed in *E. coli* cells. Transformation of pFC1-*lacZ* into *Synechocystis* 6803 and *Synechococcus* 7942 involved plasmid transfer by transconjugation: biparental mating between *E. coli* harbouring pFC1 and recipient cyanobacteria (Kreps *et al.*, 1990; Mermet-Bouvier *et al.*, 1993). Following transconjugation and antibiotic selection, heat-induction showed high expression of *lacZ* (Mermet-Bouvier and Chauvat, 1994). pFC1 has subsequently been used to allow the deletion of all chromosomal copies of the essential ferredoxin gene (*fedl*) of *Synechocystis* PCC6803. Furthermore, deletion of the *fedl* gene could be compensated by the expression of a replicating plasmid containing the *fedl* gene of the distantly related cyanobacterium *Synechococcus* PCC7942 (Poncelet *et al.*, 1998). Following this strategy, I also wanted to compensate a *ycf* 24 knockout in *Synechocystis* with over-expression of both the *E. coli* and *P. falciparum* orthologues. Unfortunately, I was unsuccessful in cloning *ycf* 24 into pFC1. However, I hope this will be done in the future.

Although the phenotypic observations I made in *Synechocystis* have given some insight into the role *ycf* 24 in the plastid of *P. falciparum* malaria parasites, further analysis of the smooth mutants could provide more substantial genetic evidence. "Pseudoreverants", transformants carrying second site suppressor mutations which restore viability under conditions that are lethal for the original mutants, can yield valuable information about other sites and/or other polypeptides that interact with the site altered by the original (primary) mutation (Yu and McIntosh, 1998; Ermakova-Gerdes and Vermaas, 1999). In the case of mutants with impaired photosynthesis, pseudorevertants with improved photosynthetic activity have been isolated and their secondary mutations analysed. One example is the pseudorevertant of a nuclear *nac1* mutant of the green alga *Chlamydomonas reinhardtii*. Such *nac1* mutants normally have reduced levels of the chloroplast-encoded D2 polypeptide of photosystem II due to defective translation of the *psbD* transcript. The pseudorevertant was also able to suppress another nuclear mutation in the *ac-115* gene which is required for synthesis of the D2 protein. The discovery of a secondary

nuclear mutation suppressing both of the other mutations suggested the participation of several gene products in the expression of a chloroplast-encoded protein (Wu and Kuchka, 1995). Another pseudorevertant, this time of *Synechocystis* sp. PCC6803, allowed suppression of mutations in the plastoquinone-binding niche of the D1 protein of photosystem II. As a consequence of these second-site mutations, mutants were restored in photosythetic function and were able to grow photoautotrophically rather than being obligate photoheterotrophs growing only on glucose. Further analysis showed tandem sequence duplications near the location of the deletions (Kleiss and Vermaas, 1995).

Mapping of suppressor mutations in an unrelated gene is usually a long, complex process. However, since the genome of *Synechocystis* PCC6803 has been sequenced the process of mapping genes that contain a second-site mutation has been simplified and accelerated by the development of a novel technique: functional complementation with size separated restriction fragment pools (Vermaas, 1998). The first step in this process is isolation of DNA from the pseudorevertant and digestion with any 5-10 of the sixteen restriction enzymes used to construct a restriction map of *Synechocystis* PCC6803. Following electrophoretic separation, the lanes of the gel are then cut into 20-25 fractions, each corresponding to a size category between ~1-20kb pairs. DNA from each fraction is then used to transform the original mutant strain. The size of the collection of restriction fragments that leads to functional complementation can then be determined for each enzyme. Comparison of the genome locations of the restriction sites and the corresponding restriction fragment sizes (available in size order from Wim Vermaas, Department of Plant Biology and Centre for the Study of Early Events in Photosynthesis, Arizona State University, USA.) should then provide a 0.5-3kb region carrying the pseudoreversion locus (Vermaas, 1998). PCR amplification of this region, either as a whole or in part, followed by transformation should lead to complementation. Sequencing of both the pseudorevertant and the primary mutant should then identify the mutation.

This technology has been used to successfully map the pseudorevertants of *psbD1* mutants of photosystem II. Two different nucleotide substitutions and a deletion of about 60% of *slr*0399 were each shown to restore photoautotrophic growth suggesting that inactivation of this open reading frame functionally complements mutations near the quinone binding site of photosystem II. As inactivation of this open reading frame in WT *Synechocystis* PCC6803 had no significant phenotypic

effects it was suggested to be a chaperone-like protein that assists, but is not essential for, quinione insertion and protein folding around Q_A in photosystem II (Ermakova-Gerdes and Vermaas, 1999).

The 16 *Synechocystis* 6803 restriction maps and size sorted restriction fragment tables have been obtained from Wim Vermaas, with a view to determining any second-site mutations which cause the pseudoreversion of ragged colony morphology to smooth. It is possible that a gene (or genes) of known function may be implicated, providing information of the true function of *ycf* 24. Conversely, the suppressor mutation(s) may be within one or more of the many ORFs of unknown function identified on the *Synechocystis* genome. However, even if their function is not known, additional gene disruption experiments could provide further phenotypic evidence for the role of *ycf* 24.

4.2 *E. coli* knockout

As with the *Synechocystis* knockout, the absence of viable *E. coli* transformants containing disrupted *ycf* 24 further implicated *ycf* 24 as an essential gene. By inference this extends to all prokaryotes carrying the gene, as well as in plastid biogenesis.

Although the results for I obtained for knockouts mediated by both linear DNA and the pKO3 gene replacement vector were as expected for an essential gene, two fundamental experiments are missing. The first of these is a positive control. Knockout of a non-essential *E. coli* gene, *ndh*, (Calhoun and Gennis, 1993; Calhoun *et al.*, 1993) was attempted using the gene replacement vector pKO3 but this failed. Whilst recombination of *ndh*Sp1-pKO3 at the *ndh* locus was implied by the formation of large colonies at the non-permissive temperature (42°C) and by PCR, selection (on plates supplemented with 5% sucrose and streptomycin) for a second cross-over event did not yield a knockout. Only non-homologous insertion of the streptomycin resistance gene was indicated, as observed for the *ycf* 24 knockout. This suggested that *ndh* is an essential gene contrary to expectation based on earlier work. However, Calhoun *et al.* (1993) used the *E. coli* host strain MWC215, and also found that in this WT strain the forced switch from NDH-2 to NDH-1, resulting from the *ndh* mutation, had less affect than expected. It was proposed that in the strain MWC215, substantial electron flux is already directed through NDH-1 and that possibly different strains

have different requirements for *ndh* genes. Perhaps in LE392 *ndh* is essential. However, a repeat experiment with pKO3, using another gene not essential to *E. coli* should be carried out. Examples of genes that could be used include the Dam methyltransferase gene, *dam* (Parker and Marinus, 1988), and that for tRNA nucleotidyl transferase, *cca*, (Zhu and Deutscher, 1987).

A second experiment that needs to be repeated is complementation of the *E. coli ycf* 24 knockout by cotransfection with *E. coli ycf* 24 carried on an expression vector. My attempt to follow this approach was unsuccessful following a knockout protocol using linear DNA introduced into *E. coli* by electroporation (Karoui *et al.*, 1997). It now needs to be attempted alongside knockout with the gene replacement vector, pKO3. Although selection of the expression vector would be required throughout the whole experiment, induction of *ycf* 24 would only be necessary following the serial dilution of integrants onto plates supplemented with 5% sucrose. Subsequent replica plating to confirm loss of vector sequence would also require induction of *ycf* 24. PCR could then be carried out to confirm inactivation at the *ycf* 24 locus. If this approach was successful it would be advantageous to attempt to complement with other *ycf* 24 orthologues, including those of *Synechocystis* and *P. falciparum*. This would test whether the *ycf* 24 orthologues are interchangeable.

An alternative method for generating a *ycf* 24 knockout in *E. coli* would be to isolate a conditional lethal mutant. Two different methods would be worth exploring. The first is similar to the episomal complementation described above, except WT *ycf* 24 would be carried on a temperature sensitive vector. Thus, at the permissive temperature replication, and therefore expression of the genes carried by the vector, would allow null *ycf* 24 mutants to be viable. Subsequent switching to the non-permissive temperature should result in cell death in the absence of vector replication and *ycf* 24 expression. However, functional complementation of disrupted *ycf* 24 was not observed in the pKO3 mediated knockout where one of the possible outcomes following knockout using this system is retention of the WT gene on pKO3, enabling lethal knockouts to be supported (Link *et al.*, 1997). Another possible way of generating temperature sensitive *ycf* 24 null mutants is random mutagenesis, as has been achieved with other genes. For example, the *fts* (filamentation thermo sensitive) genes were discovered by the isolation of conditional mutants that were unable to divide when grown under non-permissive conditions (Rothfield *et al.*, 1999). In the case of *ycf* 24, cell death would occur at the non-permissive temperature if the

knockout is lethal. An exciting consequence of either of these experiments is the possible emergence of suppressor mutants upon transfer from the permissive to the non-permissive temperature. The phenotypes associated with any secondary mutations, allowing *ycf* 24 knockout clones to remain viable even under non-permissive conditions, could provide more information on the function of *ycf* 24. Furthermore, if not associated with the *ycf* 24 locus itself, the mapping of these secondary mutations could also provide further functional evidence of the interacting gene products. However, to facilitate cloning and identification of a suppressor gene, the suppressor should have a second phenotype in addition to its suppressor effect (Appling, 1999). For example, suppressors of the temperature-sensitive *mukb* null mutant of *E. coli*, which has defective chromosome partitioning, were isolated at the non-permissive temperature. Some of these suppressors were also found to include a new phenotype, cold-sensitivity (Yamanaka *et al.*, 1992). The cold-sensitive phenotype was then exploited to clone the WT gene encoding the suppressor by complementation with a cosmid library. Suppression was thereby linked to a previously undiscovered gene *smbA*. Suppressor mutations may be mapped to genes of known function, providing direct indications of ycf 24's role, or to other ORFs of unknown function. With the latter, further mutational analysis would be required to provide functional clues.

The only indication in this project of the function of *ycf* 24 obtained from the *E. coli* knockout experiments using electroporation of linear DNA was the observation of elongated, streptomycin resistant cells. These were unlike the pUC18 control transformed cells and once more implicated a role for *ycf* 24 in cell division. However, as the cells only remained viable for a short period of time, the severe effect caused by knocking out *ycf* 24 also limited the analysis. Whilst the filamentation may have been due to induction of the SOS response, this experiment could not be repeated in a Δ*recA* host as RecA is required for homologous recombination.

4.3 Over-expression of *E. coli ycf* 24

Filamentation was observed following induction of *E. coli* Ycf 24 as an MBP fusion in the cytoplasm. This implied that the *ycf* 24 protein was either involved in cell division or induced an SOS response. The SOS response of *E. coli* is induced following exposure to agents or conditions that damage DNA or interfere in DNA replication. For example, one of the SOS genes, *sulA*, is responsible for inhibiting cell

division through interaction with FtsZ (Higashitani *et al.*, 1997). This prevents premature segregation of the damaged DNA into daughter cells during the DNA repair process (Huisman *et al.*, 1984). The two key regulatory genes of the SOS response are *recA* and *lexA* and mutation in either of these prevents induction of the SOS response. In the case of *ycf* 24, induction of the SOS response by RecA (Walker, 1996) was discounted as filamentation occurred in cells with no functional *recA*. Suppression of filamentation in a *recA*$^-$ host has previously been used to suggest that filamentation is not an effect of the SOS response. For example, Ricard and Hirota (1973) showed that a thermosensitive *ftsB* cell division mutant that forms multinucleate filaments at the restrictive temperature did not fail to filament upon mutation of *recA*. Thus the *ftsB* mutation was assumed to block septum formation. Whilst induction of pMAL-c2 and part of MSPI (merozoite surface protein I of *P. chabaudi chabaudi*) as an MBP fusion, also resulted in moderate filamentation, over-expression of *ycf* 24 with a small C-terminal His tag did not show the filamentation phenotype. These data suggest that the filamenting phenotype was attributed to induction of MBP as opposed to Ycf 24. However, the His-tagged protein was also insoluble and this may have prevented filamentation from occurring. Therefore, to determine whether Ycf 24 causes filamentation it should be expressed in soluble form from another vector. I tried a GST fusion but this also was insoluble.

Many genes, excluding those involved in the SOS response, are responsible for controlling normal cell division in *E. coli* (Rothfield *et al.*, 1999). For example, the division inhibitor MinC is required to prevent septation at unwanted sites near the cell poles. Associated with this protein is MinD, which activates MinC and MinE which topogically regulate it (Hu and Lutkenhaus, 1999). Deletions of *minC*, *minD* or all three *min* genes gives rise to a minicell phenotype, whereas inactivation of the *minE* gene alone results in the formation of long filamentous cells (Jacobs and Shapiro, 1999; Raskein and de Boer, 1997). A further nine proteins have been identified which are part of the cell division machinary of *E. coli*: FtsA, I, K, L, N, Q, W, Z and ZipA (Rothfield *et al.*, 1999). Of these, FtsZ is the most abundant and acts prior to the other proteins involved in formation of the contractile ring and cytokinesis. The action of FtsZ, prior to ZipA or any of the other *fts* proteins was deduced from the different phenotypes of the non-septate filaments of *fts* mutants. Filaments of *ftsA*, *ftsQ* and *ftsI* mutants contain broad indentations at regular intervals along the length of the cell (Rothfield *et al.*, 1999).

Using fluorescent antibody, I found that FtsZ ring formation was inhibited in cells over-expressing *ycf* 24, even though FtsZ was present in normal amounts, as shown by western blots. However, no contractile ring also was observed in cells over-expressing the MBP-αLacZ protein or part of MSPI, making a possible association between Ycf 24 over-expression and absence of FtsZ ring formation less likely. Even so, it would be interesting to look at FtsZ and other cell-division proteins in the partial (heteroplasmic) *ycf* 24 knockout of *Synechocystis* PCC6803 where genes with similarity to the *E. coli* division genes, *ftsI*, *ftsW*, *minC*, *minD* and *minE* also have been identified. As the ragged mutants tended to halt in the late stages of division, following septum formation, it seems unlikely that FtsZ utilization will be associated with these mutants.

The absence of *Synechocystis* orthologues for many of the *E. coli* genes involved in cell division implies that other genes must be present which play their roles. In particular, the absence of *zipA* indicates that more than one mechanism may exist to link the FtsZ ring to the final step of invagination (Rothfield *et al.*, 1999). Whereas both FtsZ and FtsA are peripheral membrane proteins (Tormo *et al.*, 1986), all the other *E. coli fts* division proteins appear to be integral transmembrane proteins (Rothfield *et al.*, 1999; Hale and deBoer, 1997; Dai *et al.*, 1996; Begg *et al.*, 1995). Conversely, primary sequence predictions suggest that Ycf 24 is not a transmembrane protein although it may be part of a complex with other cell division proteins or have some regulatory role.

The most interesting phenotype found on over-expression of *E. coli ycf* 24 as an MBP fusion was the appearance of aggregated nucleoids. This was not observed in any of the controls and was lost upon over-expression of a severely truncated form of *ycf* 24 (T_2 *ycf* 24). This suggests the phenotype was directly attributable to over-expression of *ycf* 24 as opposed to the over-production of MBP.

Replication of the bacterial chromosome initiates from a unique site *oriC*. At birth an *E. coli* cell contains a single copy of *oriC* and *terC*, each with a polar location. Soon after, the origins duplicate and one copy of *oriC* moves towards the opposite pole. This movement directly or indirectly causes displacement of the terminus, which then takes up a central position. Following completion of the round of replication, *terC* is duplicated at mid cell. The nucleoids are then decatenated and resolved by topological events to become two separate units. Active positioning of the

daughter chromosomes at cell quarter sites occurs before and/or during cell division. With septum formation, division occurs at mid-cell, resulting in two cells each having a copy of *terC* close to their new pole (Niki and Hiraga, 1998; Sharpe and Errington, 1999).

Mutation of a number of proteins associated with nucleoids results in defective partitioning. For example, proteins involved in decatenation of the replicated chromosome, such as DNA gyrase (encoded by *gyrA* and *gyrB*) and topisomerase IV (encoded by *parC* and *parE*), have been suggested to form a motor for rapid DNA displacement. Temperature sensitive mutants of these genes become elongated and have a large nucleoid mass in the centre of the cell at the non-permissive temperature (Kato *et al.*, 1989). In accordance with this, mutants produce more anucleate cells and many have abnormal chromosome partitioning and/or DNA content. Similar results are obtained following mutations in the two subunits of the *E. coli* DNA binding histone, HU. This protein is one of the major protein components of the bacterial nucleoid and along with the topisomerases participates in the maintenance of DNA topology (Huisman *et al.*, 1989). In a final example, reduction of FtsZ results in an increase in cell volume and retarded nucleoid segregation. Whilst I observed both of these nucleoid phenotypes in cells over-expressing *ycf* 24, the levels of FtsZ did not appear to be reduced.

In cells over-expressing *ycf* 24, chromosome replication seems to occur but nucleoids from the first round of replication appear to be laid down side by side as opposed to longitudinally in the usual way. This resembles other mutants defective in the active partitioning of daughter chromosomes which tend to replicate and decatenate normally but the nucleoids remain close to each other at the position where the chromosome was replicated (Yamanaka *et al.*, 1992). Therefore, it is possible that separation of *oriC* and/or *terC* in *E. coli* is affected by over-expression of *ycf* 24. In future work it might be possible to follow segregation of both of these genes in cells over-expressing *ycf* 24. For example, in *E. coli* and *Bacillus subtilis* insertion of multiple *lac* operator sequences into the chromosome near *oriC* allows detection of the number and position of *oriC* sequences by over-expression of a GFP-*lac* repressor fusion and observation by fluorescence microscopy (Gordon *et al.*, 1997; Webb *et al.*, 1998).

The nature of the mechanism that directs the spatially controlled separation of *oriC* is unknown. Some evidence suggests an active mitotic-like mechanism,

involving the action of a spindle-like apparatus and "motor" proteins. Several candidate genes have been identified which suggest this. One of these is the *spo0J* gene of *B. subtilis* (Ireton *et al.*, 1994), which co-localizes with *oriC* (Lewis and Errington, 1997). Mutations within this gene cause impaired chromosome partitioning with an increased proportion of anucleate cells. The production of normal-sized, chromosome-less cells is also one of the main features of *E. coli* nucleoid segregation mutants; collectively called *muk* mutants. Such mutants include *mukA* (Yamanaka *et al.*, 1996), which have mutations in the *tolC* gene encoding an outer membrane protein. However, it is not known how this protein is involved in chromosome segregation (Hiraga *et al.*, 1989). Another mutant, *mukb*, is a promising candidate for a force-generating enzyme involved in positioning the replicated and decatenated daughter chromosomes. This assumption was made because of its similarity to a microtubule associated mechanochemical enzyme, dynamin D100 (Obar, *et al.*, 1990).

Because almost any perturbation of DNA metabolism seems to lead to some degree of filamentation in *E. coli*, inhibition of septation in cells over-expressing *ycf*24 could be a mechanism to ensure co-ordination of cell division and DNA segregation (Yu and Margolin, 1999). However, after a delay many mutants with defective nucleoid segregation, resume division at some point distant from the DNA mass. This results in anucleate cells and suggests that DNA replication and segregation is not a tight, or only a temporary, regulator of cell division (Jensen and Shapiro, 1999). Anucleate cells were not observed in cells over-expressing *E. coli ycf* 24 over a period of three hours, although more detailed analysis of cells over-expressing *ycf* 24 for a longer period of time should be carried out. Unsegregated nucleoids also correlate with a lack of *ftsZ* localization (Yu and Margolin, 1999). However, unlike over-expression of *ycf* 24, in some other *par* mutants FtsZ localization is apparent in the nucleoid free gaps (Yu and Margolin, 1999).

On the other hand, the localization of MBP-Ycf 24 on either side of the condensed/unsegregated nucleoids following over-expression clearly implies some kind of association of Ycf 24 with the nucleoid. Over-production itself could be causing the abnormal phenotype, i.e. accumulation of MBP-Ycf24 around the nucleoid may prevent the active partitioning of daughter chromosomes. This accumulation was not observed in any of the controls, all of which had normal nucleoid segregation. These results, coupled with the reduced DNA content of the

partial *ycf* 24 mutants of *Synechocystis* PCC6803 and the observation of aberrant segregation of DNA in a small percentage of these cells, could indicate that *ycf* 24 affects DNA segregation. Thus cells with too little DNA (perhaps related to the levels of Ycf 24) fail to divide. In accordance with this hypothesis, the more typical DNA content observed in many cells of the smooth mutants would result in the partial resumption of cell division. However, this does not explain the blocked cytokinesis I observed at the cellular level.

4.4 Purification of *E. coli* Ycf 24

Although over-expression and purification of Ycf 24 as a soluble MBP fusion was successful, cleavage of the MBP tag with factor Xa resulted in non-specific cleavage. To address this problem I decided to clone in a different proteolytic site downstream of MBP. However, my attempt with a PreScission protease site did not succeed. Moving to other over-expression systems also did not solve this problem as both His-tagged and GST fusions were found to be insoluble. In the future I would recommend another attempt to clone a PreScission protease site into the MBP-Ycf 24 protein. This time, rather than screening potential preScission protease positive clones by sequencing as I did, large numbers could be analysed by PCR (using a primer within and outside of the site). Therefore, positive clones should be detected even if the cloning frequency was low.

The main reason for attempting to purify soluble Ycf24 was to generate monoclonal antibodies which could be used to study its localization in host cells, perhaps using immuno-EM. This might have indicated cellular associations under normal growth conditions, giving information on the function of the *ycf* 24 protein. An antibody to Ycf24 would also confirm the peri-nucleoid localization of MBP-Ycf24 in over-expressing cells. So far immunization with the MBP fusion protein (Malcolm Strath, NIMR) has only resulted in antibodies against the MBP tag. Therefore, soluble Ycf24 minus the MBP tag is required to repeat the immunization.

The biological functions of many proteins depend on binding interactions with other molecules, including other proteins (Vergnon and Chu, 1999). For example, protein-protein complexes are critical for the speed and precision of signal transduction (Pawson and Scott, 1997). Also MucB, MucC and MucF form a complex involved in chromosome partitioning in *E. coli* (Yamazoe *et al.*, 1999). Another principle reason for producing both purified Ycf24 and antibodies to it was to carry

out protein-protein interaction studies. As Ycf 24 is of unknown function any interactions with other proteins and/or complex formation could provide information on its cellular role. But although purified Ycf24 and its corresponding antibody are desirable, they may not be essential to the study of protein-protein interactions. Both immunoprecipitation and affinity chromatography can be used with purified, soluble fusion protein and antibodies to the tag.

Immunoprecipitation is one of the classical methods for detecting protein-protein interactions. The basic principle involves the generation of cell lysates to which antibody is added. The antigen is then precipitated and washed before bound proteins are eluted and analysed (Phizicky and Fields, 1995). A commercially available antibody for MBP could be used in such an experiment; e.g. radiolabelled *E. coli* proteins (produced in cells incubated with a radiolabelled amino acid such as S ^{35}-methionine) are solubilized by lysis to release labelled proteins that are immunoreactive, undegraded and biologically active. Subsequent incubation with unlabelled purified fusion protein (as a carrier) and an antibody against the fusion tag should result in the formation of specific immune complexes. These can then be bound to the appropriate affinity beads (amylose, in the case of MBP) before collection and elution. Analysis of the immunoprecipitate can be done by SDS-PAGE separation and autoradiography. Any radiolabelled proteins identified might then be sequenced by mass spectrophometry (Sults, 1990; Fenselau *et al.*, 1993, Yates, 2000) or by N-terminal sequencing (Tsugita, 1987; Hoving *et al.*, 2000). Immunoprecipitation using fusion proteins is widely used. For example, a GST fusion has been used to characterise a GTP-dependent interaction between Ras and Raf (Warne *et al.*, 1993). Affinity chromatography may also be useful in the context of Ycf24 as it utilises fusion proteins with purification tags, such as MBP- and His-tags (Beeckmans, 1999). The former can be immobilized by amylose and the latter by iminodiacetate or nitriloacetate charged with various metal ions in a column. Interacting proteins from a bacterial cell lysate which bind to the immobilized fusion can then be eluted, analysed by SDS-PAGE and sequenced.

Relatively recently the yeast two-hybrid system has been developed which allows the screening of large numbers of potential protein interactions and does not require purified protein as a ligand or an antibody for detection (Fields and Song, 1989; Chien *et al.*, 1991; Fields and Sternglauz, 1994). The underlying principle behind this technique is the co-transfection of a host yeast strain with plasmids

containing two in-frame gene fusions. The first of these plasmids includes coding sequences for the Gal4 DNA binding domain with the coding sequence of a protein of interest (the bait). The second is a gene fusion of coding sequences for the Gal4 transcription activation domain with the coding sequence for a test protein, or with an assortment of coding sequences derived from a genomic or cDNA library (the prey). Expression of the chimeric proteins is achieved using the constitutive promoter of the yeast *ADH1* gene, and nuclear localization is directed by internal sequences in the binding domain or SV-40 sequence tagged on to the amino terminus of the activation domain. The bait hybrid is able to bind by itself to an operator adjacent to the promoter. If the fusion proteins interact, the bait hybrid bound to the operator will recruit the prey hybrid from solution and position it close to the operator, activating it. The interaction between the two can be made to generate an easily detectable phenotype by placing a suitable reporter gene under the control of the prey-dependent promoter. An alternative to the Gla4 system is the LexA two-hybrid system (Golemis and Brent, 1992), which employs a viral activation domain, VP16, and the *E. coli* LexA DNA-binding domain. As for the Gal4 system, positive interactions are scored by activation of reporter genes with promoters containing operator sequences (this time LexA operator sequences).

The *Saccharomyces cerevisiae* two-hybrid system has been used successfully to show protein-protein interaction between *E. coli* proteins. For example, using this system Jonczyk and Nowicka (1996) showed interaction between the UmuD and UmuC SOS mutagenesis proteins. A future prospect therefore, is to screen an *E. coli* library using this method to identify proteins which interact with Ycf24.

However, before embarking on a screen it is important to determine that the bait protein of interest is expressed, and is stable and properly folded in yeast. Although expression can be verified by immunoblotting (either with antibodies against the prey protein or the Gal4 binding domain) functional assays to show stability and proper folding are difficult if no interaction partners are already known. Furthermore, once putative interaction partners have been identified it has been suggested that true interactions need to be verified by a second experimental method (McAlister-Henn *et al.*, 1999).

Alternative methods for studying protein-protein interactions include co-immunoprecipitation and affinity chromatography as mentioned already.

More recently, an *E. coli* two-hybrid system has been developed which overcomes some of the problems associated with the yeast based system (Hu *et al.*, 2000). These improvements include better transformation efficiency, fewer false positives and negatives, less toxicity, better library coverage as well as no requiremen t for localization. These advantages, as well as its speed make it a good alternative method for attempting protein-protein interactions with *ycf* 24 if verification of stable expression and correct folding in yeast cannot be achieved.

4.5 Overall conclusion

To address the function of the malarial plastid gene *ycf* 24 I disrupted its orthologue in two bacterial surrogate systems, *Synechocystis* PCC6803 and *E. coli*. The failure to achieve a true knockout in either organism indicates that *ycf* 24 is an essential gene. This inference can probably be extended to all other prokaryotes carrying the gene as well as to plastid biogenesis in plants and in *Plasmodium falciparum*. The perturbations of bacterial cell division and DNA segregation observed following the partial knockout of *ycf* 24 in *Synechocystis*, and its over-expression in *E. coli*, suggest that *ycf* 24 is required for cytokinesis and/or DNA segregation. These conclusions are not immediately apparent from *ycf* 24's annotation in the database where it is described sometimes as a subunit of an ATP-transporter, and no mechanistic link between the phenotype and such a function have been found. Further experiments are required to clarify this discrepency.

Attempts were also made to purify *E. co*li Ycf 24 for antibody production and other studies. Although purification of a fusion protein was successful, cleavage of Ycf 24 from its MBP tag failed. Immunization of mice with the fusion protein has still to generate an antibody to Ycf 24. However, it is hoped that both of these goals will be achieved in the near future since this would assist with the study of Ycf 24 protein-protein interactions and provide more insight into the cellular role of this essential plastid organelle gene product.

Appendices

Appendix 1. Compositions of general solutions and buffers.

All solutions made up in dH_2O unless otherwise stated.

AGAROSE GEL ELECROPHORESIS GEL LOADING BUFFER

0.25% (w/v)	bromophenol blue
0.25% (w/v)	xylene cyanol FF
3% (v/v)	glycerol

COOMASSIE STAIN

45% (v/v)	methanol
10% (v/v)	ethanoic acid
0.1% (w/v)	Coomassie Brilliant Blue R

CTAB-NaCl

700mM	NaCl
10% (v/v)	CTAB

DESTAIN

25% (v/v)	propan-2-ol
10% 9v/v)	ethanoic acid
3% v/v)	glycerol

HYBRIDIZATION SOLUTION

362.5mM	NaCl
20mM	Sodium phosphate buffer (pH7.7)
1.9875mM	EDTA
1% (w/v)	SDS
0.5% (w/v)	skimmed milk powder (Marvel)
10% (w/v)	dextran sulphate
250μg/ml	denatured salmon sperm DNA

NATIVE LOADING BUFFER (1x)

50mM	Tris
2mM	PMSF
1mM	EDTA
10% (v/v)	glycerol
0.1% (w/v)	bromophenol blue
pH7.5	

PBS (1x)

0.1M	NaCl
80mM	Na_2HPO_4
20mM	NaH_2PO_4

SALINE-EDTA BUFFER

0.15M	NaCl
0.01M	EDTA
pH8.0	

SDS LOADING BUFFER (1x)

1% (w/v)	SDS
50mM	Tris
2mM	PMSF
1mM	EDTA
10% (v/v)	glycerol
0.5% (v/v)	β-2-mercaptoethanol

SOUTHERN WASH 1

2x	SSPE
0.1% (w/v)	SDS

SOUTHERN WASH 2

1x	SSPE
0.1% (w/v)	SDS

SOUTHERN WASH 3

0.1x	SSPE
0.1% (w/v)	SDS

SSC (1x)

0.15M	NaCl
0.015M	tri-sodium citrate

SSPE (1x)

120mM	NaCl
15mM	tri-sodium citrate
13mM	KH_2PO_4
1mM	EDTA

TAE BUFFER (1x)

900mM	Tris-acetate
2mM	EDTA
pH8.0	

TBE BUFFER (1x)

10.8g/l	Tris
22g/l	Boric acid
0.75g/l	EDTA
pH8.0	

TE BUFFER

10mM	Tris
1mM	EDTA
pH8.0	

TES BUFFER (1x)

5mM	Tris
50mM	NaCl
5mM	EDTA

Appendix 2. Isolation of genomic DNA from *Synechocystis* PCC6803.

(Peter Nixon, Imperial College, London)

Aim: To isolate small quantities of DNA for Southern blotting.

1. Centrifuge 12ml of dense culture (OD at A_{750} of at least 2.0) or scrape a pea sized glob of cells from a plate of growing cells. If the OD is less than 2.0, increase the volume of cells used. Spin down for at least 15 minutes at 4,000rpm.
2. Resuspend in 400µl TES.
3. Make a solution of lysozyme (50mg/ml in distilled water) and add 100µl. Incubate for 15 minutes at 37°C, mixing occasionally to prevent cells from settling out.
4. Add 50µl 10% sarkosyl, then 600µl phenol and vortex about once a minute for ten minutes.
5. Spin in microfuge (13,000rpm) for at least 5 minutes.
6. Dilute 10mg/ml RNase stock x 20 in sterile distilled water and add 5µl.
7. Incubate for 15 minutes at 37°C.
8. Add 100µl of 5M NaCl, 100µl of CTAB-NaCl solution and 600µl of chloroform. Vortex about once a minute for 10 minutes.
9. Spin for 2 minutes (13,000rpm). Transfer supernatant to new tubes and precipitate with 700µl of isopropanol. Spin for 15 minutes (as before) to pellet.
10. Rinse pellet with 70% EtOH, dry under a vacuum or air-dry. Resuspend in 100µl TE.
11. Store at 4°C or -20°C long term.

Appendix 3A. Primer sequences.

Organism	Name and Tm.	Direction	Sequence
Synechocystis PCC6803	A3, 65°C	5′	CTC CGC CCC TAA GCA AAG TAA GAA A
	A5, 65°C	3′	TGA TCC TCG CCA ATT TTG GAA GTG GA
	A6, 68°C	5′	GAG CTC GGG ATC CCG ATG AGT TCC ACC ACT GTT A
	A7, 68°C	3′	GAG CTC GGA TCC TTA ACC CAC AGT ACC TTC TA
	P1, 50°C	5′	AAT TGT GAG CGG ATA ACA
	P2, 52°C	3′	GTT GGG TAA CGC CAG GGT
	K1, 50°C	5′	TGA CAC CAA TCA ACT TCA
	K2, 50°C	3′	CAG CGT ACC AGT TTT GTA
E. coli	*ycf*24-5P, 62°C	5′	GAG CTC GGA ATT CGC ATG TGG CTG TGG CGA AAG
	*ycf*24-3P, 62°C	3′	GAG CTC GGG ATC CTT ATC CGA CGC TGT GTT CAA G
	A8, 55°C	5′	GAC GAC ACT TGC GCG TCT
	A9, 55°C	3′	CAC GGG AAC GCC CAA CTG
	A11, 65°C	5′	GAG CTC GGA ATT CAT GTG GCT GTG GCG AAA G
	A9*, 65°C	3′	GAG CTC GCT GCA GTT ATC CGA CGC TGT GTT CAA
	pET28a1 65°C	5′	GAG CTC CAT GGG ATG GCT GTG GCG AAA G
	pET28a2 65°C	3′	GAG CTC CTC GAG TCC GAC GCT GTG TTC A
	A13, 45°C	5′	TCG TCT AAA CGC CTA TC
	A14, 45°C	5′	TGA AGG CTG TTC CGC
	A15, 45°C	5′	TCT CTG CCG GAC ATA G
	A16, 65°C	3′	GAG CTC CTG CAG TTA ATA AGT AGT GGC AAC CGA AAC TGA
	A17, 65°C	3′	GAG CTC CTG CAG TTA CGG TTC TTC CAT CTC CAG
	A18, 65°C	5′	GAG CTC GGA TCC ATG TGG CTG TGG CGA AAG
	A19, 65°C	3′	GAG CTC GAA TTC TTA TCC GAC GCT GTG TTC

E. coli	6P1-1	5′	CAG CTG GAA TTC CTG GAA GTT CTG TTC CAG GGG CCC GAA TTC CAG CTG
	6P1-2	3′	CAG CTG GAA TTC GGG CCC CTG GAA CAG AAC TTC CAG GAA TTC CAG CTG
	APK1, 68°C	5′	GAG CTC GGA TCC ATG TGG CTG TGG CGA AAG CTT TGG
	APK2, 68°C	3′	GAG CTC GGA TCC TTA TCC GAC GCT GTG TTC AAG
	PK03-L, 55°C	5′	AGG GCA GGG TCG TTA AAT AGC
	PKO3-R, 58°C	3′	TTA ATG CGC CGC TAC AGG GCG
	A20, 50°C	5′	TCA GAT ATT CAA ATT TCA
	A21, 50°C	3′	TGT AAA TCT TTA ATA CTT
	APK6, 65°C	3′	GAG CTC GGA TCC TTA ATG CAA CTT CAA ACG
	APK7. 65°C	5′	GAG CTC GGA TCC TTG ACT ACG CCA TTG
	APK12, 50°C	5′	GAG CAG TTG GGC GTT
	APK13, 50°C	3′	TAC GCA CCA GTT CCG
	APK14, 50°C	5′	TTC AAT ACG CCA GGT GTC
	APK15, 50°C	3′	AGA GTT CTA CAC CCG TCG
	APK16, 68°C	5′	GAG CTC GGA TCC ATG GAC ATG CAT TCA GGA ACC TTT AAC CCA CAA G
	APK17, 68°C	3′	GAG CTC GGA TCC TTA CTG CTG TTC GGT AAG CCA GCC ATA ACC CAG C
	APK18, 55°C	5′	GTT CCA CCA CCA GCT GGT TG
	APK19, 50°C	3′	CAC GCT TGG TGA CGA AC
	APK20, 52°C	5′	GCT GTA ACC TGT TGT TAA TTA
	APK21, 52°C	3′	ATG CCA CAT CCG CCA GTG TAC

Appendix 3B. Primer definitions

***Synechocystis*:**

A3 (fwd)	PCR 1kb fragment of *ycf*24 for cloning into pUC (knockout).
A5 (rev)	
A6 (fwd)	PCR entire *ycf*24 for cloning into pBSHdSp1, *Bam* HI (rescue)
A7 (rev)	*Bam* HI
P1 (fwd)	Sequencing 1kb *ycf*24 in pUC9
P2 (rev)	
K1 (fwd)	Sequencing kanamycin resistance gene in *ycf*24
K2 (rev)	

***E. coli*:**

ycf24-3p (fwd)	PCR *ycf*24, *Eco* RI.
ycf24-5p (rev)	*Bam* HI.
A8 (fwd)	Sequencing Sp1 (streptomycin resistance cassette) in *ycf*24.
A9 (rev)	
A11 (fwd)	Cloning *ycf*24 into pMAL vectors in frame, *Eco* RI.
A9* (rev)	*Pst* I.
pET28a1 (fwd)	Cloning *E.coli ycf*24 into pET28a. In frame with the ATG site in the *Nco* I cloning site.
pET28a2 (rev)	Cloning *E.coli ycf*24 into pET28a, *Xho* I. Also in frame so the His- tag is transcribed at the 3′ end. For this purpose the *ycf*24 stop codon is excluded from the primer sequence.
A13 (fwd)	Internal *E. coli* primer for sequencing gene. Position 216-232.
A14 (fwd)	Internal *E. coli* primer for sequencing gene. Position 792-806.
A15 (fwd)	Internal primer for sequencing gene. Position 1169-1184.
A16 (rev)	Internal primer for cloning the 5′ 500bp of *ycf*24 into pMAL-c2. In frame and including a *Pst* I site and stop codon. Position 481-504.
A17 (rev)	Internal primer for cloning the 5′ 250bp of *ycf*24 into pMAL-c2. In frame and including a *Pst* I site and stop codon. Position 241-258.
A18 (fwd)	Cloning *ycf*24 into pGEX-6P1, *Bam* HI.

A19 (rev)	*Eco* RI.
6P1-1 (fwd)	Cloning PreScission proteolytic site from pGEX-6P1 into *ycf* 24-pMAL-c2, *Eco* RI.
6P1-2 (rev)	*Eco* RI.
APK1 (fwd)	Cloning *ycf*24Sp1 into pKO3, *Bam* HI.
APK2 (rev)	*Bam* HI.
PKO3-L (fwd)	Sequencing primer for pKO3 (Link *et al.*, 1997).
PKO3-R (rev)	
A20 (fwd)	PCR primer for checking recombination into *E. coli* genome - just 5′ of *ycf*24.
A21 (rev)	PCR primer for checking recombination into *E. coli* genome - just 3′ of *ycf*24.
APK6 (rev)	Cloning *ndh* gene into pKO3, *Bam* HI.
APK7 (fwd)	*Bam* HI.
APK12 (fwd)	Sequencing Sp1 in *ycf*24 - either side of *Nru* I cloning site.
APK13 (rev)	
APK14 (fwd)	Sequencing Sp1 in *ndh* gene - either side of *Xmn* I cloning site.
APK15 (rev)	
APK16 (fwd)	PCR either side of *ycf*24 and its 2 flanking genes after recombination of the vector + insert into the *E. coli* genome.
APK17 (rev)	
APK18 (fwd)	Internal sequencing primer for *ndh* gene (position 878-897).
APK19 (rev)	Internal sequencing primer for *ycf*24 (position 936-952).
APK20 (Fwd)	PCR primer for checking recombination into *E. coli* genome - just 5′ of *ndh.*
APK 21 (rev)	PCR primer for checking recombination into *E. coli* genome - just 5′ of *ndh.*

Appendix 4. Staining with Reynolds lead citrate for EM

(Liz Hirst, NIMR)

1. 1.33g of $Pb(NO_3)_2$ lead nitrate and 1.76g of $Na_3C_6H_5O_7 \cdot H_20$ tri-sodium citrate each in 15ml of sterile dH_2O added together and shaken for 1 minute.
2. Leave for 30 minutes to react, shaking intermittently.
3. Add 8mls of 1M carbonate-free NaOH and mix by inversion until the solution clears.
4. Make up to 50mls with sterile dH_2O.
5. Store in an airtight container (with KOH), in the cool and dark (the stain lasts for approximately six weeks).
6. To stain place a drop of lead citrate on parafilm (American National Can.™) covered dental wax (taking care not to breathe on it) in a Petri-dish (Sterilin) containing a few pellets of NaOH.
7. Float the grids (containing the specimens) on the drop of lead citrate specimen side down.
8. Wash with dH_2O. Holding the grid with forceps and dropping the water over it using a syringe.

Appendix 5. Genomic locations of *ycf*24 and *ycf*16 (database survey to 15/4/00)

Prokaryotes:

Organism	*ycf24* (gene name) (location)	*ycf16* (gene name) (location)	Source
Aeropyrum pernix	APE1703 88222-89652	APE1702 87392-88153	*A. pernix* genomic DNA (Accession: AP000062)
Archaeoglobus fulgidus	AF2365 8551-9660	AF2364 7829-8593	*A. fulgidus* complete genome (Accession: AE001113)
Methanobacterium thermoautotrophicum	MTH1150 9887-11119	MTH1149 9150-9905	*M. thermoautotrophicum* complete chromosome (Accession: AE000884)
Methanococcus jannaschii	?	MJ0035 complement (3053-3805)	*M. jannaschii* complete genome (Accession: U67462)
Pyrococcus abyssi	PAB1856 complement (117814-119154)	PAB1855 complement (119147-119890)	*P. abyssi* complete genome (Accession: AJ248285)
Pyrococcus furiosus	ORF4 2481-3908	ORF3 1748-2491	(Accession: AF156097)
Pyrococcus horikoshii	PH1385 83216-84556	PH1384 82480-83223	*P. horikoshii* OT3 genomic DNA (Accession: AP000006)
Bacillus subtilis	*yurU* complement (157544-159049)	*yurY* complement (162054-162839)*	*B. subtilis* complete genome (Accession: Z99120)
Chlamydia pneumoniae	CPn0692 complement (6439-7893)	abcX complement (5663-6433)	*C. pneumoniae* complete genome (Accession: AE001651)
Chlamydia trachomatis	CT684 203-1654	CT685 1657-2424	*C. trachomatis* complete genome (Accession: AE001339)
Deinococcus radiodurans	DR2106 complement (454-1860)	DR2107 complement (1955-2713)	*D. radiodurans* complete chromosome (Accession: AE002046)
Escherichia coli	*ynhE* complement (5394-6920)	*ynhD* complement (4638-5384)	*E. coli* K-12 complete genome (Accession: AE000263)
Mycobacterium leprae	MLCL536.28c complement (20425-23034)	MLCL536.26c complement (18471-19253)*	*M. leprae* cosmid L536 (Accession: Z99125)
Mycobacterium tuberculosis	Rv1461 7343-9883	Rv1463 11070-11870*	*M. tuberculosis* H37Rv complete genome

			(Accession: AL021184)
Rickettsia prowazekii	?	RP508 30678-31442	*R. prowazekii* strain Madrid E, complete genome (Accession: AJ235272)
Shigella dysenteriae	unknown <2669->2979	?	*S. dysenteriae pinD* gene for part of phage tail fiber, site-specific recombinase phage Mu Com-like protein and phage Mu Mom-like protein (Accession: D63765)
Streptomyces coelicolor	SCC22.07c complement (4299-5720)	SCC22.04c complement (1968-2732)*	*S. coelicolor* cosmid C22 (Accession: AL096839)
Thermotoga maritima	TM1369 829-2223	TM1368 92-832	*T. maritima* complete genome (Accession: AE001791)
Treponema pallidum	TP0612 4854-6293	TP0611 3994-4773	Complete genome sequence of *T. pallidum* (Accession: AE001236)
Anabaena	Yes	Yes	*Anabaena* sequencing project, Cyanobase (contig c292)
Synechocystis	slr0074 2908-4350	slr0075 4492-5262	*Synechocystis* sp. PCC6803 complete genome (Accession: D64004)

* 1 or more genes in between.

Protoctista:

Organism	*ycf24* (gene name) (location)	*ycf16* (gene name) (location)	Source
Guillardia theta	61067-62518	62544-63305	*G. theta* complete plastid genome (Accession: AF041468)
Antithamnion sp.	*orf3* (partial) complement (5672->6565)	*orf2* complement (4893-5648)	Chloroplast large ATP synthase operon (Accession: X63382)
Cyanidium caldarium (= *Galdieria sulphuraria*)	*ycf24* 114386-115837	*ycf16* 115838-116602	*C. caldarium* RK1 chloroplast sequence (Accession: AF022186
Cyanophora paradoxa	complement (67893-69353)	complement (67117-67896)	*C. paradoxa* cyanelle, complete genome (Accesssion: U30821)
Porphyra purpurea	40948-42411	42408-43163	*P. purpurea* chloroplast, complete genome (Accession: U38804)
Odontella sinensis	69086-70546	70546-71301	*O. sinensis* complete chloroplast genome (Accession: Z67753)
Skeletonema costatum	chromoplast partial *ycf24* gene (Accession: AJ132267)	?	Plastid DNA sequences of *S. costatum* NIES 323
Eimeria tenella	436->1597	?	*E. tenella* plastid DNA (Accession: Y12333)
Plasmodium berghei	ORF470 complement (<1-711)	?	*P. berghei* extrachromosomal plastid PB-1, partial (Accession: U79731)
Plasmodium chabaudi	ORF473 62-1483	?	*P. chabaudi* unknown plastid genes (Accession: AF200327)
Plasmodium falciparum	ORF470 5142-6554 *P. falciparum* complete gene map of plastid-like DNA (IR-A) (Accession: X95275)	Chromosome 14?	
Toxoplasma gondii	ORF470 complement (28289-29686)	Nucleus-encoded?	*T. gondii* chloroplast, complete genome (Accession: U87145)

Plants:

Organism	***ycf24*** (gene name) (location)	***ycf16*** (gene name) (location)	**Source**
Arabidopsis thaliana	AT4g04770 complement (join (155014-155232, 155347-156801)) (Chromosome 4 contig fragment No. 13) (Accession: AL161501)	T7M13.25? <79104..>80335 *A. thaliana* BAC T7M13 (Chromosome 3) (Accession: AC011708)	
Glycine max	?	EST partial (Accession: AI442209)	Public Soybean EST Project
Lycopersicon esculentum	?	EST partial (Accession: AW032235)	*L. esculentum* cDNA clone cLEC35L20

Bibliography

Aaij, C., and Borst, P. (1972) The gel electrophoresis of DNA. Biochim. Biophys. Acta *269*, 192-200.

Aikawa, M. (1966). The fine structure of the erythrocytic stage of three avian malarial parasites *Plasmodium fallax*, *P. lophurae*, and *P. cathemerium*. Am. J. Trop. Med. Hyg. *15*, 449-471.

Aikawa, M., Huff, C. G., and Sprinz, H. (1967). Fine structure of the asexual stages of *Plasmodium elongatum*. J. Cell Biol. *34*, 229-249.

Allen, J. F., and Raven, J. A. (1996). Free-radical-induced mutation versus redox regulation: costs and benefits of genes in organelles. J. Mol. Evol. *42*, 482-492.

Allen, M. M. (1984). Cyanobacterial cell inclusions. Ann. Rev. Microbiol. *38*, 1-25.

Appling, D. R. (1999). Genetic approaches to the study of protein-protein interactions. Methods *19*, 338-349.

Apt, K. E., and Grossman, A. R. (1993). Characterization and transcript analysis of the major phycobiliprotein subunit gene from *Algaothamnion neglectum* (Rhodopyta). Plant Mol. Biol. *21*, 27-38.

Baldauf, S. L., Manhart, J. R., and Palmer, J. D. (1990). Different fates of the chloroplast *tufA* gene following transfer to the nucleus in green algae. Proc. Natl. Acad. Sci. (USA) *87*, 5317-5321.

Barta, J. R., Jenkins, M. C., and Danforth, H. D. (1991). Evolutionary relationship of avian *Eimeria* species among other apicomplexan protozoa: monophyly of the apicomplexa is supported. Mol. Biol. Evol. *8*, 345-355.

Bartling, D., Clausmeyer, S., Oelmuller, R., and Herrmann, R. G. (1990). Towards epitope models for chloroplast transit sequences. Bot. Mag. (Tokyo) *Special Issue 2*, 119-144.

Beekmans, S. (1999). Chromatographic methods to study protein-protein interactions. Methods *19*, 278-305.

Begg, K. J., Dewar, S., and Donachie, W. (1995). A new *Escherichia coli* cell division gene *ftsK*. J. Bacteriol. *177*, 6211-6222.

Berzins, K., and Perlmann, P. (1996). In: Malaria vaccine develpment and a multi stage perspective, S. L. Hoffman, ed. (Washington, DC: Am. Soc. Microbiol.), pp. 105-143.

Blanchard, J. L., and Hicks, J. S. (1999). The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. J. Euk. Microbiol. *46*, 367-375.

Bloomfield, I. C., Vaughn, V., Rest, R. F., and Eisenstein, B. I. (1991). Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature sensitive PSC101 replicon. Mol. Microbiol. *5*, 1447-1457.

Bodyl, A. (1997). Mechanism of protein targeting to the chlorarachniophyte plastids and the evolution of complex plastids with four membranes - a hypothesis. Bot. Acta. *101*, 395-400.

Bohne, W., Gross, U., Ferguson, D. J., and Heesemann, J. (1995). Cloning and characterization of a bradyzoite-specifically expressed gene (*hsp30/bag1*) of *Toxoplasma gondii*, related to genes encoding small heat-shock proteins of plants. Mol. Microbiol. *16*, 1221-1230.

Borst, P., Overdulve, J. P., Weijers, P. J., Fase-Fowler, F., and Van Den Berg, M. (1984). DNA circles with cruciforms from *Isospora* (*Toxoplasma*) *gondii*. Biochim. Biophys. Acta. *781*, 100-111.

Boussiba, S., and Richmond, A. E. (1980). C-phycocyanin as a storage protein in the blue/green alga *Spirulina platensis*. Arch. Microbiol. *125*, 143-147.

Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., Mclean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C., and Barrell, B. G. (1999). The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. Nature *400*, 532-538.

Boyer, C., Shannon, J. C., and Hardison, R. C. (1989). Physiology, biochemistry, and genetics of non-green plastids, Volume 379 (Rockville: American Society of Plant Physiologists).

Brown, E. D, Vivas, E. L., Walsh, C. T. (1995). MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis is essential in *Escherichia coli*. J. Bacteriol. *177*, 4194-4197.

Bryant, D. A. (1987). The cyanobacterial photosynthetic apparatus: comparisons to those of higher plants and photosynthetic bacteria. Can. Bull. Fish. Aquat. Sci. *214*, 423-501.

Bryant, D. A., and Tandeau de Marsac, N. (1988). Isolation of genes encoding components of photosynthetic apparatus. Methods Enzymol. *167*, 755-765.

Burnette, W. N. (1981). Western blotting: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radiodinated protein A. Anal. Biochem. *112*, 195-203.

Calhoun, M. W., and Gennis, R. B. (1993). Demonstration of separate genetic loci encoding distinct membrane-bound respiratory NADH dehydrogenases in *Escherichia coli*. J. Bacteriol. *175*, 3013-3019.

Calhoun, M. W., Oden, K. L., Gennis, R. B., de Mattos, M. J., and Neijssel, O. M. (1993). Energetic efficiency of *Escherichia coli*: effects of mutations in components of the aerobic respiratory chain. J. Bacteriol. *175*, 3020-3025.

Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. Microbiol. Rev. *57*, 953-994.

Cavalier-Smith, T. (1995). Zooflagellate phylogeny and classification. Tsitologica *37*, 1010-1029.

Cedergren, R., Gray, M. W., Abel, Y., and Sankoff, D. (1988). The evolutionary relationships among known life forms. J. Mol. Evol. *28*, 98-112.

Cermakian, N., Ikeda, T. M., Cedergren, R., and Gray, M. W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucl. Acids Res. *24*, 648-654.

Cerutti, H., and Jagendorf, A. (1995). Movement of DNA across the chloroplast envelope: implications for the transfer of promiscuous DNA. Photosyn. Res. *46*, 329-337.

Cerutti, H., Johnson, A. M. Boynton, J. E., and Gilham, N. W. (1995). Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Escherichia coli recA*. Mol. Cell. Biol. *15*, 3003-3011.

Chan, R., Keller, M., Canaday, S., Wiel, J., and Imbault, P. (1990). Eight small subunits of *Euglena* ribulose 1-5 bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein. EMBO J. *9*, 333-338.

Chien, C., Bartel, P. L., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc. Natl. Acad. Sci. (USA) *88*, 9578-9582.

Chitnis, P. R., Purvis, D., and Nelson, N. (1991). Molecular cloning and targeted mutagenesis of the gene *psaF* encoding subunit III of photosystem I from the cyanobacterium *Synechocystis* sp. PCC6803. J. Biol. Chem. *266*, 20146-20151.

Chitnis, P. R., Reilly, P. A., and Nelson, N. (1989). Insertional inactivation of the gene encoding subunit II of photosystem I from the cyanobacterium *Synechocystis* sp. PCC 6803. J. Biol. Chem. *264*, 18381-18385.

Claesson, C., Lustig, F., Boren, T., Simonsson, C., Barciszewska, M., and Lagerkvist, U. (1995). Gycine codon discrimination and the nucleotide in position 32 of the anticodon loop. J. Mol. Biol. *274*, 191-196.

Clark, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucl. Acids Res. *16*, 9677-9686.

Cline, K., Fulsom, D. R., and Viitanen, P. V. (1989). An imported thylakoid protein accumulates in the stroma when insertion into thylakoid membranes is inhibited. J. Biol. Chem. *264*, 14225-14232.

Cline, K., and Henry, R. (1996). Import and routing of nucleus-encoded chloroplast proteins. Ann. Rev. Cell Dev. Biol. *12*, 1-26.

Cline, K., Werner-Washburne, M., Lubben, T. H., and Keegstra, K. (1985). Precursors to two nuclear-encoded chloroplast proteins bind to the outer envelope membrane before being imported into chloroplasts. J. Biol. Chem. *260*, 3691-3696.

Clough, B., Rangachari, K., Strath, M., Preiser, P. R., and Wilson, R. J. M. (1999). Antibiotic inhibitors of organellar protein synthesis in *Plasmodium falciparum*. Protist *150*, 189-195.

Clough, B., Strath, M., Preiser, P., Denny, P., and Wilson, R. J. M. (1997). Thiostrepton binds to malarial plastid rRNA. FEBS Lett. *406*, 123-125.

Clyde, D. F., Most, H., McCarthy, V. C., and Vanderberg, J. P. (1973). Immunization of man against sporozoite-induced *falciparum* malaria. Am. J. Med. Sci. *266*, 169-177.

Cohen, G. B., Ren, R., and Baltimore, D. (1995). Molecular binding domains in signal transduction proteins. Cell *80*, 237-248.

Cohen, S., McGregor, I. A., and Carrington, S. C. (1961). Gamma-globulin and acquired immunity to human malaria. Nature *192*, 733-737.

Collier, J. L., and Grossman, A. R. (1992). Chlorosis induced by nutrient deprivation in *Synechococcus* sp. strain PCC7942: not all bleaching is the same. J. Bacteriol. *174*, 4718-4726.

Colston, M. J., and Davis, E. O. (1994). The ins and outs of protein splicing elements. Mol. Microbiol. *12*, 359-363.

Creasey, A., Mendis, A., Carlton, J., Williamson, D., Wilson, I., and Carter, R. (1994). Maternal inheritance of extrachromosomal DNA in malaria parasites. Mol. Biochem. Parasitol. *65*, 95-98.

Cronan, J. E., and Rock, C. O. (1996). Biosynthesis of membrane lipids. In: *Escherichia coli* and *Salmonella*: cellular and molecular biology, F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Rezenikoff, M. Riley, M. Schaechter and H. E. Umbarger, eds. (Washington, DC: American Society for Microbiology), pp. 612-636.

Croxden, F. E., and Cowden, D. J. (1962). The Chi-square test. In: Applied general statistics (London: Sir Issac Pitman and Sons Ltd.), pp. 681-693.

Cundliffe, E. (1986). Thiostrepton and the ribosomal GTPase centre. In: Structure, function and genetics of ribosomes, B. Hardesty and G. Kramer, eds.: Springer Verlag), pp. 590.

Dabert, P., and Smith, G. R. (1997). Gene replacement with linear DNA fragments in wild-type *Escherichia coli*: enhancement by chi sites. Genetics *145*, 877-889.

Dai, K., Xu, Y., and Lutkenhaus, J. (1996). Topological characterization of the essential *Escherichia coli* cell division protein FtsN. J. Bacteriol. *175*, 1328-1334.

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

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

de Lorimer, R., Bryant, D. A., Porter, R. D., Liu, W., Jay, E., and Stevens, S. E. J. (1984). Genes for the α and β subunits of phycocyanin. Proc. Natl. Acad. Sci. (USA) *81*, 7946-7950.

Dedonder, R. (1966). Levansucrase from *Bacillus subtilis*. Methods Enzymol. *8*, 500-505.

Denny, P., Williamson, D., and Wilson, I. (1998). Evidence for a single origin of the 35kb plastid DNA in apicomplexans. Protist *149*, 51-59.

Devilly, C. I., and Houghton, J. A. (1977). A study of genetic transformation in *Gleoeocapsa alpicola*. J. Gen. Microbiol. *98*, 277-280.

Dietzel, I., Klob, V., and Boos, W. (1978). Pole cap formation in *Escherichia coli* following induction of the maltose binding protein. Arch. Microbiol. *118*, 207-208.

Dixon, D. A., and Kowalczykowski, S. C. (1993). The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. Cell *73*, 87-96.

Dobberstein, B., Blobel, G., and Chua, N. H. (1977). *In vitro* synthesis and processing of a putative precursor for the small subunit of ribulose-1, 5-bisphosphate carboxylase of *Chlamydomonas reinhardii*. Proc. Natl. Acad. Sci. (USA) *74*, 1082-1085.

Ellis, J. (1982). Promiscous DNA: chloroplast genes inside plant mitochondria. Nature *299*, 678-679.

Emlyn-Jones, D., Ashby, M. K., and Mullineaux, C. W. (1999). A gene required for the regulation of photosynthetic light-harvesting in the cyanobacterium *Synechocystis* 6803. Mol. Microbiol. *33*, 1050-1058.

Ermakova-Gerdes, S., and Vermaas, W. (1999). Inactivation of the open reading frame *slr*0399 in *Synechosystis* sp. PCC6803 functionally complements mutations near the Q_A niche of photosystem II. J. Biol. Chem. *274*, 30540-30549.

Escalante, A. A., and Ayala, F. J. (1995). Evolutionary origin of *Plasmodium* and other apicomplexa based on rRNA genes. Proc. Natl. Acad. Sci. (USA) *92*, 5793-5797.

Fawcett, T. W., and Bartlett, G. (1990). An effective method for eliminating "artefact banding" when sequencing double-stranded DNA templates. Biotechniques *9*, 46-48.

Feagin, J. E. (1992). The 6kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. Mol. Biochem. Parasitol. *52*, 1455-148.

Feagin, J. E., and Drew, M. E. (1995). *Plasmodium falciparum*: alterations in organelle transcript abundance during the eythrocytic cycle. Exp. Parasitol. *80*, 430-440.

Feagin, J. E., Werner, E., Gardner, M. J., Williamson, D. H., and Wilson, R. J. (1992). Homologies between the contiguous and fragmented rRNAs of the *Plasmodium falciparum* extrachromosomal DNAs are limited to core sequences. Nucl. Acids Res. *20*, 879-887.

Fenselau, C., Vestling, M. M., and Cotter, R. J. (1993). Mass spectrometric analysis of proteins. Curr. Opin. Biotechnol. *4*, 14-19.

Fields, S., and Song, O.-K. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Fields, S., and Sternglanz, R. (1994). The two-hybrid system: An assay for protein:protein interactions. Trends Genet. *10*, 286-292.

Fischer, N., Stampacchia, O., Redding, K., and Rochaix, J. D. (1996). Selectable marker recycling in the chloroplast. Mol. Gen. Genet. *251*, 373-380.

Fischera, M. E., and Roos, D. S. (1997). A plastid organelle as a drug target in apicomplexan parasites. Nature *390*, 407-409.

Flores, E., and Herrero, A. (1994). Assimilatory nitrogen metabolism and its regulation, In: The molecular biology of cyanobacteria, D. A. Bryant, ed. (The Netherlands: Kluwer Academic Publishers), pp. 487-517.

Flores, E., and Muro-Pastor, A. M. (1990). Mutational and kinetic analysis of basic amino acid transport in the cyanobacterium *Synechocystis* sp. PCC6803. Arch. Microbiol. *154*, 521-527.

Friedman, A. L., and Keegstra, K. (1989). Chloroplast protein import. Quantitative analysis of precursor binding. Plant Physiol. *89*, 993-999.

Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V., and Sogin, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates and ciliates. Mol. Biochem. Parasitol. *45*, 147-154.

Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H., and Waring, M. J. (1981). Antibiotic inhibitors of ribosome function. In: The molecular basis of antibiotic action, E. F. Gale, E. Cundliffe, P. E. Reynolds, M. H. Richmond and M. J. Waring, eds. (London: John Wiley and Sons Ltd.), pp. 403-545.

Gantt, J. S., Baldauf, S. L., Calie, P. J., Weeden, N. F., and Palmer, J. D. (1991). Transfer of *rpl22* to the nucleus greatly preceeded its loss from the chloroplast and involved the gain of an intron. EMBO J. *10*, 3073-3075.

Gardner, M. J., Bates, P. A., Ling, I. T., Moore, D. J., McCready, S., Gunasekera, M. B. R., Wilson, R. J. M., and Williamson, D. H. (1988). Mitochondrial DNA of the human malarial parasite *Plasmodium falciparum*. Mol. Biochem. Parasitol. *31*, 11-18.

Gardner, M. J., Feagin, J. E., Moore, D. J., Rangachari, K., Williamson, D. H., and Wilson, R. J. M. (1993). Sequence and organization of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite *Plasmodium falciparum*. Nucl. Acids Res. *21*, 1067-1071.

Gardner, M. J., Feagin, J. E., Moore, D. J., Spencer, D. F., Gray, M. W., Williamson, D. H., and Wilson, R. J. M. (1991*a*). Organisation and expression of small subunit ribosomal RNA genes encoded by a 35-kilobase circular DNA in *Plasmodium falciparum*. Mol. Biochem. Parasitol. *48*, 77-88.

Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D. C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G. G., Clayton, R., White, O., Smith, H. O., Fraser, C. M., Adams, M. D., Venter, J. C., and Hoffman, S. L. (1998). Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. Science *282*, 1126-1132.

Gardner, M. J., Williamson, D. H., and Wilson, R. J. M. (1991*b*). A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. Mol. Biochem. Parasitol. *44*, 115-124.

Gay, P., Le CoQ, M., Steinmetz, M., Berkelman, T., and Kado, C. I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. *164*, 918-921.

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

Gimble, F. S., and Thorner, J. (1992). Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. Nature *357*, 301-306.

Goldberg, A. L. (1992). The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. Eur. J. Biochem. *203*, 9-23.

Golden, S. S. (1988). Mutagenesis of cyanobacteria by classical and gene transfer-based methods. Methods Enzymol. *167*, 714-727.

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

Golemis, E. A., and Brent, R. A. (1992). Fused protein domains inhibit DNA binding by LexA. Mol. Cell. Biol. *12*, 3006-3014.

Gordon, S., Sitinikov, D., Webb, C. D., Teleman, A., Losick, R., Murray, A. W., and Wright, A. (1997). Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinctive mechanisms. Cell *90*, 1113-1121.

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

Grigorieva, G., and Shestakov, S. (1982). Transformation in the cyanobacterium *Synechocystis* sp. 6803. FEMS Microbiol. Lett. *13*, 367-370.

Grossman, A. R., Schaefer, M. R., Chiang, G. G., and Collier, J. L. (1993). The phycobilisome, a light harvesting complex response to environmental conditions. Microbiol. Rev. *57*, 725-749.

Grossman, A. R., Schaefer, M. R., Chiang, G. G., and Collier, J. L. (1994). The responses of cyanobacteria to environmental conditions: light and nutrients. In: The Molecular Biology of Cyanobacteria, D. A. Bryant, ed. (The Netherlands: Kluwer Academic Publishers), pp. 641-675.

Gutteridge, W. E., Trigg, P. I., and Williamson, D. H. (1971). Properties of DNA from some malarial parasites. Parasitology *62*, 209-219.

Hackstein, J. H., Mackenstedt, U., Mehlhorn, H., Meijerink, J. P., Schubert, H., and Leunissen, J. A. (1995). Parasitic apicomplexans harbor a chlorophyll a-D1 complex, the potential target for therapeutic triazines. Parasitol. Res. *81*, 207-216.

Hale, C., and de Boer, P. (1997). Direct binding of FtsZ to ZipA, an essential component of the septal ring structure that mediates cell division in *E. coli*. Cell *88*, 175-185.

Halpin, C., Musgrove, J. E., Lord, J. M., and Robinson, C. (1989). Import and processing of proteins by castor bean leucoplasts. FEBS Lett. *258*, 32-34.

Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P., and Kushner, S. R. (1989). New method for generating deletions and gene replacements in *Escherichia coli*. J. Bacteriol. *171*, 4617-4622.

Harder, A., and Habercorn, A. (1989). Possible mode of action of toltrazuril: studies on two *Eimeria* species and mammalian and *Ascaris suum* enzymes. Parasitol. Res. *76*, 8-12.

Hayat, M. A. (1978). Introduction to biological scanning E.M. (London: University Park Press).

Henze, K., Badr, A., Wettern, M., Cerff, R., and Martin, W. (1995). A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. (USA) *92*, 9122-9126.

Hernandez-Torres, J., Breitenberger, C. A., Spielmann, A., and Stutz, E. (1993). Cloning and sequencing of a soybean nuclear gene coding for a chloroplast translation elongation factor EF-G. Biochim. Biophys. Acta. *1174*, 191-194.

Herrmann, R. G. (1997). Eukaryotism, towards a new interpretation. In: Eukaryotism and symbiosis, H. E. A. Schenk, R. G. Herrmann, K. W. Jeon, N. E. Muller and W. Schwemmler, eds. (Heidelberg: Springer), pp. 73-118.

Higashitani, A., Ishii, Y., Kato, Y., and Horiuchi, K. (1997). Functional dissection of a cell-division inhibitor, SulA, of *Escherichia coli* and its negative regulation by Lon. Mol. Gen. Genet . *254*, 351-357.

Higgins, C. F. (1995). The ABC of channel regulation. Cell *82*, 693-696.

Hiraga, S., Niki, H., Ogura, T., Ichinose, C., Mori, H., Ezaki, B., and Jaffe, A. (1989). Chromosome partitioning in E. coli: novel mutants producing anucleate cells. J. Bacteriol. *171*, 1496-1505.

Hiraga, S., Niki, H., Ogura, T., Ichinose, C., Mori, H., Ezaki, B., and Jaffe, A. (1989). Chromosome partitioning in *Escherichia coli* novel mutants producing anucleate cells. J. Bacteriol. *171*, 1496-1505.

Hirsh, D. (1971). Tryptophan transfer RNA as the UGA supressor. J. Mol. Biol. *58*, 439-458.

Hirsh, D., and Gold, L. (1971). The translocation of the UGA triplet *in vitro* by tryptophan transferase RNAs. J. Mol. Biol. *58*, 459-468.

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

Holder, A. A. (1999). Malaria vaccines. Proc. Natl. Acad. Sci. (USA) *96*, 1167-1169.

Hollingshead, S., and Vapneck, D. (1985). Nucleotide sequence analysis of a gene encoding a streptomycin/spectinomycin adenyltransferase. Plasmid *13*, 17-30.

Hopkins, J., Fowler, R., Krishna, S., Wilson, I., Mitchell, G., and Bannister, L. (1999). The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis. Protist *150*, 283-295.

Houlne, G., and Schantz, R. (1988). Characterization of cDNA sequences for LHCI apoproteins of *Euglena gracilis*: the mRNA encodes a large precursor containing several consecutive divergent polypeptides. Mol. Gen. Genet. *213*, 479-486.

Hoving, S., Munchbach, M., Schmid, H., Signor, L., Lehmann, A., Staudenmann, W., Quadroni, M., and James, P. (2000). A method for the chemical generation of N-terminal sequence tags for rapid protein identification. Anal. Chem. *72*, 1006-1014.

Howe, C. J. (1992). Plastid origin of an extrachromosomal DNA molecule from *Plasmodium*, the causative agent of malaria. J. Theoret. Biol. *158*, 199-205.

Howe, C. J., and Smith, A. G. (1991). Plants without chlorophyll. Nature *349*, 109.

Howells, R. E., and Fullard, F. J. (1970). Mitochondrial changes during the malarial life cycle. Trans. Royal Soc. Trop. Med. Hyg. *64*, 6.

Hu, J. C., Kornacker, M. G., and Hochschild, A. (2000). *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein-protein interactions. Methods *20*, 80-94.

Hu, Z., and Lutkenhaus, J. (1999). Topological regulation of cell division in *Escherichia coli* involves rapid pole to pole oscillation of the division inhibitor MinC under the control of MinD and MinE. Mol. Microbiol. *34*, 82-90.

Huisman, O., D'Ari, R., and Gottesman, S. (1984). An inducible DNA replication-cell division coupling mechanism in *E. coli*. Nature *290*, 797-799.

Huisman, O., Faelen, M., Girard, D., Jaffe, A., Toussaint, A., and Rouviere-Yaniv, J. (1989). Multiple defects in *Escherichia coli* mutants lacking HU protein. J. Bacteriol. *171*, 3704-3712.

Huls, P. G., Vischer, N. O. E., and Woldringh, C. L. (1999). Delayed nucleoid segregation in *Escherichia coli*. Mol. Microbiol. *33*, 959-970.

Hyde, J. E., Kelly, S. L., Holloway, S. P., Snewin, V. A., and Sims, P. F. G. (1989). A general approach to isolating *Plasmodium falciparum* genes using non-redundant oligonucleotides inferred from protein sequences of other organisms. Mol. Biochem. Parasitol. *32*, 247-262.

Ireton, K., Gunther, N. W., and Grossman, A. D. (1994). *spo0J* is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. J Bacteriol *176*, 5320-5329.

Jacobs, C., and Shapiro, L. (1999). Bacterial cell division: a moveable feast. Proc. Natl. Acad. Sci. (USA) *96*, 5891-5893.

Jemiolo, D. K., Pagel, F. T., and Murgola, E. J. (1995). UGA supression by a mutant RNA of the large ribosomal subunit. Proc. Natl. Acad. Sci. (USA) *92*, 12309-12313.

Jensen, R. B., and Shapiro, L. (1999). Chromosome segregation during the prokaryotic cell division cycle. Curr. Opin. Cell Biol. *11*, 726-731.

Jiang, X.-M., Brahmbhatt, H. N., Quigley, N. B., and Reeves, P. R. (1987). A low copy number cosmid. Plasmid *18*, 170-172.

Jonczyk, P., and Norwicka, A. (1996). Specific *in vivo* protein-protein interactions between *Escherichia coli* SOS mutagenisis proteins. J. Bacteriol. *178*, 2580-2585.

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H9(+)-adenosine triphosphate. Science *250*, 651-657.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. *3*, 109-136.

Karoui, M. E., Amunsden, S. K., Dabert, P., and Gruss, A. (1999). Gene replacement with linear DNA in electroporated wild-type *Escherichia coli*. Nucl. Acids Res. *27*, 1296-1299.

Kato, J., Nishimura, Y., and Suzuki, H. (1989). *Escherichia coli parA* is an allele of the *gryB* gene. Mol. Gen. Genet. *217*, 178-181.

Keegstra, K. (1989). Transport and routing proteins into chloroplasts. Cell *56*, 247-253.

Keegstra, K., and Cline, K. (1999). Protein import and routing systems of chloroplasts. Plant Cell *11*, 557-570.

Kerber, B., and Soll, J. (1992). Transfer of chloroplast-bound precursor protein into the translocation apparatus is impaired after phospholipase C treatment. FEBS Lett. *306*, 71-74.

Kiatfuengfoo, R., Suthiphongchai, T., Prapunwattana, P., and Yuthavong, Y. (1989). Mitochondria as the site of action of tetracycline on *Plasmodium falciparum*. Mol. Biochem. Parasitol. *34*, 109-116.

Kilejian, A. (1975). Circular mitochondrial DNA from the avian malarial parasite *Plasmodium lophurae*. Biochim. Biophys. Acta *390*, 276-284.

Kirwin, P. M., Elderfield, P. D., and Robinson, C. (1987). Transport of proteins into chloroplasts. Partial purification of a thylakoidal processing peptidase involved in plastocyanin biogenesis. J. Biol. Chem. *262*, 16386-16390.
Kishore, G. M., and Shah, D. M. (1988). Amino acid biosynthesis inhibitors as herbicides. Ann. Rev. Biochem. *57*, 627-663.

Kleiss, H., and Vermass, W. (1995). Tandem sequence duplications functionally complement deletions in the D1 protein of photosystem II. J. Biol. Chem. *270*, 16536-16541.

Knell, A. J. (1991). Malaria (Oxford: Oxford University Press).

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

Kostrzewa, M., and Zetsche, K. (1992). Large ATP synthase operon of the red alga *Antithamnion* sp. resembles the corresponding operon in cyanobacteria. J. Mol. Biol. *227*, 961-970.

Kreps, S., Ferino, F., Mosrin, C., Gerits, J., Mergeay, M., and Thuriaux, P. (1990). Conjugative transfer and autonomous replication of a promiscous IncQ plasmid in the cyanobacterium *Synechocystis* PCC6803. Mol. Gen. Genet. *221*, 129-133.

Kuroiwa, T., Kuroiwa, H., Saki, A., Takahashi, H., Toda, K., and Itoh, R. (1998). The division apparatus of plastids and mitochondria. Int. Rev. Cytol. *191*, 1-41.

Labarre, J., Chauvat, F., and Thuriaux, P. (1989). Insertional mutagenesis by random cloning of antibiotic reistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC6803. J. Bacteriol. *171*, 3449-3457.

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

Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M. W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in minature. Nature *387*, 493-497.

Lang-Unnasch, N., and Aiello, D. P. (1999). Sequence evidence for an altered genetic code in the *Neospora caninum* plastid. Int. J. Parasitol. *29*, 1557-1562.

Lea, P. J., Wills, W. R., Wallsgrove, R. M., and Miflin, B. J. (1982). Assimilation of nitrogen and synthesis of amino acids in chloroplasts and cyanobacteria (blue-green algae). In: On the Origins of Chloroplasts, J. A. Schiff, ed. (North Holland: Elserver), pp. 149-178.

Levine, N. D. (1987). Phylum II. Apicomplexa. In: An illustrated guide to the protozoa. Society of Protozoologists, J. J. Lee, H. Hunter and E. C. Bovee, eds. (Kansas: Lawrence), pp. 322-357.

Levine, N. D. (1988). Progress in taxonomy of the apicomplexan protozoa. J. Protozool. *35*, 518-520.

Lewis, P. J., and Errington, J. (1997). Direct evience for active segregation of *oriC* regions of the *Bacillus subtilis* chromosome and co-localization with the SpoOJ partitioning protein. Mol. Microbiol. *25*, 945-954.

Lewis, R. J., Tsai, F. T., and Wigley, D. B. (1996). Molecular mechanisms of drug inhibition of DNA gyrase. BioEssays *18*, 661-671.

Li, S. J., and Cronan, J. E. J. (1992). Putitave zinc finger protein encoded by a conserved chloroplast gene is very likely a subunit of a biotin dependent carboxylase. Plant Mol. Biol. *20*, 759-761.

Link, A. J., Phillips, D., and Church, G. M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. J. Bacteriol. *179*, 6228-6237.

Liu, X. Q., and Hu, Z. (1997). Identification and characterization of a cyanobacterial DnaX intein. FEBS Lett. *408*, 311-314.

Lonsdale, D. M., Hodge, T. P., and Fauron, C. M. (1984). The physical map and organisation of the mitochondrial genome from the fertile cytoplasm of maize. Nucl. Acids Res. *12*, 9249-9261.

Lutkenhaus, J. (1998). Organelle division: From *coli* to chloroplasts. Curr. Biol. *8*, 619-621.

Lynch, M. (1996). Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. *13*, 209-220.

MacPherson, G. G., Warrell, M. J., Looareesuwan, S., and Warrell, D. A. (1985). Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. Am. J. Pathol. *119*, 385-401.

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. *3*, 208-218.

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

Martin, W., and Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. Nature *392*, 37-41.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., and Hasegawa, M. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. Nature *393*, 162-165.

Mass, M. J., and Roop, B. C. (1992). PCR primers specific for detection of a rat repetitive sequence. Biotechniques *13*, 676-678.

Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., and Gottesman, S. (1990*a*). Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli*. J. Biol. Chem. *265*, 12536-12545.

Maurizi, M. R., Clark, W. P., Kim, S.-H., and Gottesman, S. (1990*b*). ClpP represents a unique family of serine proteases. J. Biol. Chem. *265*, 12546-12552.

McAlister-Henn, L., Gibson, N., and Panisko, E. (1999). Applications of the yeast two-hybrid system. Methods *19*, 330-337.

McEntee, K. (1977). Genetic analysis of the *Escherichia coli* K-12 *slr* region. J. Bacteriol. *132*, 904-911.

McFadden, G. I., Reith, M. E., Munholland, J., and Lang-Unnasch, N. (1996). Plastid in human parasites. Nature *381*, 482.

McFadden, G. I., and Waller, R. F. (1997). Plastids in parasites of humans. BioEssays *19*, 1033-1040.

McFadden, G. I., Waller, R. F., Reith, M. E., and Lang-Unnasch, N. (1997). Plastids in apicomplexan parasites. Plant Systemat. Evol. *S11*, 261-287.

McGregor, I. A. (1960). Demographic effects of malaria with special reference to the stable malaria of Africa. W. African Med. J. *6*, 260-265.

Mehlhorn, H., Ortmann-Faukstein, E., and Haberkorn, A. (1984). The effects of sym. triazones on the developmental stages of *Emeria tenella*, *E. maxima* and *E. acervulina*. A light and electron microscopial study. J. Parasitol. *70*, 173-182.

Mermet-Bouvier, P., Cassier-Chauvat, C., Marraccini, P., and Chauvat, F. (1993). Transfer and replication of RSF1010-derived plasmids in several cyanobacteria of the genera *Synechocystis* and *Synechococcus*. Curr. Microbiol. *27*, 323-327.

Mermet-Bouvier, P., and Chauvat, F. (1994). A conditional expression vector for the cyanobacteria *Synechocystis* sp. strains PCC6803 and PCC6714 or *Synechococcus* sp. strains PCC7942 and PCC6301. Curr. Microbiol. *28*, 145-148.

Miller, L. H. (1997). The spirit of Dakar. A call for action on malaria. Nature *386*, 541.

Mita, T., Kanbe, T., Kanaka, K., and Kuroiwa, T. (1986). A ring structure around the dividing plane of the *Cyanidium caldarium* chloroplast. Protoplasma *130*, 211-213.

Mitchell, G. H., Butcher, G. A., and Cohen, S. (1975). Merozoite vaccination against *Plasmodium knowlesi* malaria. Immunology *29*, 397-407.

Mitra, S. K., Lustig, F., Akesson, B., and Lagerkvist, U. (1977). Codon-anticodon recognition in the valine codon family. J. Biol. Chem. *252*, 471-478.

Miziorko, H. M., and Lorimer, G. H. (1983). Ribulose-1,5-bisphosphate carboxylase-oxygenase. Ann. Rev. Biochem. *52*, 507-535.

Muller, H. J. (1964). The relation of recombination to mutational advance. Mut. Res. *1*, 2-9.

Murgola, E. J., Hijazi, K. A., Goringer, H. U., and Dahlberg, A. E. (1988). Mutant 16S ribosomal RNA: a codon translocation supressor. Proc. Natl. Acad. Sci. (USA) *85*, 4162-4165.

Murphy, K. C. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J. Bacteriol. *180*, 2063-2071.

Nass, M. M. K., and Nass, S. (2000). Fibrous structures within the matrix of the developing chick embryo mitochondria. Exp. Cell Res. *255*, 1-3.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. *10*, 1-6.

Nierhaus, K. H. (1996). An elongation factor turn-on. Nature *379*, 491-492.

Niki, H., and Hiraga, S. (1998). Polar localization of the replication origin and terminus in *Escherichia coli* nucleoids during chromosome partitioning. Genes Devel. *12*, 1036-1045.

Nugent, J. M., and Palmer, J. D. (1991). RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell *66*, 473-481.

Obar, R. A., Collins, C. A., Hammarback, J. A., Shpetner, H. S., and Vallee, R. B. (1990). Molecular cloning of the microtubule-associated mechanochemical enzyme dynamin reveals homology with a new family of GTP-binding proteins. Nature *347*, 256-261.

Osawa, S., Jukes, T. H., Watanabe, K., and Muto, A. (1992). Recent evidence for the evolution of the genetic code. Microbiol. Rev. *56*, 229-264.

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. Ann. Rev. Genet. *19*, 325-354.

Palmer, J. D. (1990). Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. *6*, 115-120.

Palmer, J. D., and Delwiche, C. F. (1996). Second hand chloroplasts and the case of the disappearing nucleus. Proc. Natl. Acad. Sci. (USA) *93*, 7432-7435.

Palmer, J. D., and Shields, C. R. (1984). Tripartate structure of the *Brassica campestris* mitochondrial genome. Nature *307*, 437-440.

Parker, B., and Marinus, M. G. (1988). A simple and rapid method to obtain substitution mutations in *Escherichia coli*: isolation of a *dam* deletion/insertion mutation. Gene *73*, 531-535.

Pawson, T., and Scott, J. D. (1997). Signalling through scaffold, anchoring, and adaptor proteins. Science *278*, 2075-2080.

Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucere, R. B., Benner, J., Slatko, B. E., Nwankwo, D. O., and Hempstead, S. K. (1992). Intervening sequences in an Archaea DNA polymerase gene. Proc. Natl. Acad. Sci. (USA) *89*, 5577-5581.

Phizcky, E. M., and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. Microbiol. Rev. *59*, 94-123.

Poncelet, M., Cassier-Chauvat, C., Leschelle, X., Bottin, H., and Chauvat, F. (1998). Targeted deletion and mutational analysis of the essential (2Fe-2S) plant-like ferredoxin in *Synechocystis* PCC6803 by plasmid shuffling. Mol. Microbiol. *28*, 813-821.

Pont-Kingdon, G. A., Okada, N. A., Macfarlane, J. L., Beaglet, C. T., Wolstenholme, D. R., Cavalier-Smith, T., and Clark-Walker, G. D. (1995). A coral mitochondrial *mutS* gene. Nature *375*, 109-111.

Porter, R. D. (1986). Transformation in cyanobacteria. CRC Crit. Rev. Microbiol. *13*, 111-131.

Preiser, P., Williamson, D. H., and Wilson, R. J. M. (1995). tRNA genes transcribed from the plastid-like DNA of *Plasmodium falciparum*. Nucl. Acids Res. *23*, 4329-4336.

Preiser, P. R., Wilson, R. J. M., Moore, P. W., McCready, S., Hajibagheri, M. A. N., Blight, K. J., Strath, M., and Williamson, D. H. (1996). Recombination associated with replication of the malarial mitochondrial DNA. EMBO J. *15*, 684-693.

Quentin, Y., Fichant, G., and Denizot, F. (1999). Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. J. Mol. Biol. *287*, 467-484.

Raskein, and de Boer, P. (1997). The MinE ring: an FtsZ-independent cell structure required for selection of the correct division site in *E. coli*. Cell *91*, 685-694.

Read, M., Hicks, K. E., Simms, P. F., and Hyde, J. E. (1994). Molecular characterisation of the enolase gene from the human malaria parasite *Plasmodium falciparum*. Evidence for ancestry within a photosynthetic lineage. Eur. J. Biochem. *220*, 513-520.

Reithman, H. C., Mawhinney, T. P., and Sherman, L. A. (1988). Characterization of phycobilisome glycoproteins in the cyanobacterium *Anacystis nidulans* R2. J. Bacteriol. *170*, 2433-2440.

Ricard, M., and Hirota, Y. (1973). Process of cellular division in *Escherichia coli*: physiological study on thermosensitive mutants defective in cell division. J. Bacteriol. *116*, 314-322.

Rijdersberg, C. P., Amez, J., Thielen, P. G. M., and Swager, J. A. (1979). Fluorescence emission spectra of chloroplast and subchloroplast preparations at low temperature. Biochim. Biophys. Acta. *545*, 473-482.

Rijgesberg, C. P., and Amez, J. (1980). Fluorescence and energy transfer in phycobiliprotein-containing algae at low temperature. Biochim. Biophys. Acta. *593*, 261-271.

Rippka, R., Derulles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. *111*, 1-61.

Roberts, F., Roberts, C. W., Johnson, J. J., Kyle, D. E., Krell, T., Coggins, J. R., Coombs, G. H., Milhous, W. K., Tzipori, S., Ferguson, D. J. P., Chakrabarti, D., and McLeod, R. (1998). Evidence for the shikimate pathway in apicomplexan parasites. Nature *393*, 801-805.

Robson, K. J., Gamble, Y., and Acharya, K. R. (1993). Molecular modelling of malaria calmodulin suggests that it is not a suitable target for novel anti-malarials. Phil. Trans. Royal Soc. London Ser. B *340*, 39-53.

Rothfield, L., Justice, S., and Garcia-Lara, J. (1999). Bacterial cell division. Ann. Rev. Genet. *33*, 423-448.

Roy, A., Cox, R. A., Williamson, D. H., and Wilson, R. J. (1999). Protein synthesis in the plastid of *Plasmodium falciparum*. Protist *150*, 183-188.

Rudzinska, M. A., and Trager, W. (1968). The fine structure of trophozoites and gametocytes in *Plasmodium coatneyi*. J. Protozool. *15*, 73-88.

Rudzinska, M. A., and Vickerman, K. (1968). Intracellular phagotrophy in *Babesia rodhaini* as revealed by electron microscopy. J. Protozool. *9*, 279-288.

Rush, M. G., and Misra, R. (1985). Extrachromosomal DNA in eukaryotes. Plasmid *14*, 177-191.

Russell, C. B., Thaler, D. S., and Dahlquist, F. W. (1989). Chromosomal transformation of *Escherichia coli recD* strains with linearized plasmids. J. Bacteriol. *171*, 2609-2613.

Sakai, R. K., Gelford, D. H., Stoeffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487-491.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, USA: Cold Spring Harbor Press).

Sanger, F., Niklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. (USA) *74*, 5463-5467.

Schindler, C., and Soll, J. (1986). Protein import in intact, purified pea etioplasts. Arch. Biochem. Biophys. *274*, 211-220.

Schuster, W., and Brennicke, A. (1987). Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera:* is genetic information transferred between organelles via RNA. EMBO J. *6*, 2857-2863.

Shanklin, J., DeWitt, N. D., and Flanagan, J. M. (1995). The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease. Plant Cell *7*, 1713-1722.

Sharpe, M. E., and Errington, E. (1999). Upheaval in the bacterial nucleoid. Trends Genet. *15*, 70-74.

Sherman, D. M., and Sherman, L. A. (1983). Effect of iron deficiency and iron restoration on the ultrastructure of *Anacystis nidulans*. J. Bacteriol. *156*, 393-401.

Shestakov, S. V., and Khuyen, N. T. (1970). Evidence for genetic transformation in the blue-green alga *Anacystis nidulans*. Mol. Gen. Genet. *107*, 372-375.

Shevell, D. E., Zamzam, A., Demple, A. M., and Walker, G. C. (1988). Construction of an *Escherichia coli* K-12 *ada* deletion by gene replacement in a *redD* strain reveals a second methyltransferase that repairs alkylated DNA. J. Bacteriol. *170*, 3294-3296.

Shibata, T., DasGupta, C., Cunningham, R. P., and Radding, C. M. (1979). Purified *Escherichia coli recA* protein catalyses homologous pairing of superhelical DNA and single-stranded fragments. Proc. Natl. Acad. Sci. (USA) *76*, 1638-1642.

Silberman, J. D., Sogin, M. L., Leipe, D. D., and Clark, C. G. (1996). Human parasite finds taxonomic home. Nature *380*, 398.

Simon, R. D. (1987). Inclusion bodies in the cyanobacteria: cyanophycin, polyphosphate, polyhedral bodies. In: The cyanobacteria, P. Fay and C. van Baalen, eds. (Amsterdam: Elsevier Biomedical Publishers), pp. 199-225.

Small, I., Suffolk, R., and Leaver, C. J. (1989). Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell *58*, 69-76.

Smeekens, S., Weisbeek, P., and Robson, C. (1990). Protein transport into and within chloroplasts. Trends Biochem. Sci. *15*, 73-76.

Smeijsters, L. J. J. W., Zijlstra, N. M., de Vries, E., Franssen, F. J., Jansa, C. J., and Overdulve, J. P. (1994). The effect of (S)-9-(3-hydroxy-2-phosphonyl-methoxypropyl) adenine on nuclear and organellar DNA synthesis in erythrocytic shizogony in malaria. Mol. Biochem. Parasitol. *67*, 115-124.

Smith, S. M., and Ellis, R. J. (1979). Processing of small subunit precursor of ribulose biphosphate carboxylase and its assembly into whole enzyme are stromal events. Nature *278*, 662-664.

Soll, J., and Alefsen, H. (1993). The protein import apparatus of chloroplasts. Physiologia Plantarum *87*, 433-440.

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

Stanier, G. (1988). Fine structure of cyanobacteria. Methods Enzymol. *167*, 157-172.

Steinberg, S., Misch, A., and Sprinzl, M. (1993). Compilation of tRNA sequences and sequences of tRNA genes. Nucl. Acids Res. *21*, 3011-5.

Stern, D. B., and Lonsdale, D. M. (1982). Mitochondrial and chloroplast genomes of maize have a 12kb DNA sequence in common. Nature *299*, 698-702.

Stevens, S. E., and Porter, R. D. (1980). Transformation in *Agmenellum quadruplicatum*. Proc. Natl. Acad. Sci. (USA) *77*, 6052-6056.

Stevens, S. E. J., Paone, D. A. M., and Balkwill, D. L. (1981). Accumulation of cyanophycin granules as a result of phosphate limitation in *Agmenellum quadruplicatum*. Plant Physiol. *67*, 716-719.

Stewart, G. J., and Carlson, C. A. (1986). The biology of natural transformation. Ann. Rev. Microbiol.*40*, 211-135.

Stoute, J. A., Kester, K. E., Krzych, U., Wellde, B. T., Hall, T., White, K., Glenn, G., Okenhouse, C. F., Garcon, N., Schwenk, R., Lanar, D. E., Sun, P., Momin, P., Wirtz, R. A., Golenda, C., Slaoui, M., Wortmann, G., Holland, C., Dowler, M., Cohen, J., and Ballou, W. R. (1998). Long term efficacy and immune responses following immunization with the RTS,S malaria vaccine. J. Inf. Dis. *178*, 1139-1144.

Sugiura, M. (1992). The chloroplast genome. Plant. Mol. Biol. *19*, 149-168.

Sults, J. (1990). Peptide sequencing by mass spectrometry. Methods Biochem. Anal. *34*, 145-201.

Surolia, N., and Padmanaban, G. (1992). *De novo* biosynthesis of heme offers a new chemotherapeutic target in the human malarial parasite. Biochem. Biophys. Res. Comm. *187*, 744-750.

Taylor, F. J. R. (1974). Flagellate phylogeny: a study in conflicts. Taxonomy *23*, 229-258.

Thatcher, T. H., and Gorovsky, M. A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3 and H4. Nucl. Acids. Res. *22*, 174-179.

Thornber, J. P. (1986). Biochemical characterization and structure of pigment-proteins of photosynthetic organisms. Encl. Plant Physiol. *19*, 98-142.

Tomii, K., and Kanehisha, M. (1998). A comparative analysis of ABC transporters in complete microbial genomes. Genome Res. *8*, 1048-1059.

Tormo, A., Ayala, J. A., de Pedro, M. A., Aldera, M., and Vicente, M. (1986). Interaction of FtsA and PBP3 proteins in the *Escherichia coli* septum. J. Bacteriol.. *166*, 985-992.

Tsugita, A. (1987). Developments in protein microsequencing. Adv. Biophys. *23*, 81-133.

Vaidya, A. B., Akella, R., and Suplick, K. (1989). Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. Mol. Biochem. Parasitol. *35*, 97-108.

Vaidya, A. B., Morrisey, J., Plowe, C. V., Kaslow, D. C., and Wellems, T. E. (1993). Unidirectional dominance of cytoplasmic inheritance in two genetic crosses of *Plasmodium falciparum*. Mol. Cell. Biol. *13*, 7349-7357.

Van de Peer, Y., Van der Auwera, G., and De Wachter, R. (1996). The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. J. Mol. Evol. *42*, 201-210.

Vergnon, A. L., and Chu, Y.-H. (1999). Electrophoretic methods for studying protein-protein interactions. Methods *19*, 270-277.

Vermaas, W. F. J. (1998). Gene modifications and mutation mapping to study the function of photosystem II. Methods. Enzymol. *297*, 293-310.

von Heijne, G., and Nishikawa, K. (1991). Hypothesis: chloroplast transit peptides. The perfect random coil? FEBS Lett. *278*, 1-3.

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

Walker, G. C. (1996). The SOS response of *Escherichia coli*. In: *Escherichia coli* and *Salmonella*: cellular and molecular biology, F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Rezenikoff, M. Riley, M. Schaechter and H. E. Umbarger, eds. (Washington, DC: American Society for Microbiology), pp. 1400-1416.

Waller, R. F., Keeling, P. J., Donald, R. G. K., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., and McFadden, G. I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc. Natl. Acad. Sci. (USA) *95*, 12352-12357.

Waller, R. F., Reed, M. B., Cowman, A. F., and McFadden, G. I. (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. EMBO J. *19*, 1794-1802.

Wallsgrove, R. M. (1991). Plastid genes and parasitic plants. Nature *350*, 664.

Warne, P. H., Viciana, P. R., and Downward, J. (1993). Direct interaction of Ras and the amino-terminal region of Raf-1 *in vitro*. Nature *364*, 352-355.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M. (1988). The molecular biology of the gene, 4th Edition, J. R. Gillen, ed. (Calfornia: The Benjamin/Cummins Publishing company Inc.).

Wealand, J. L., Myers, J. A., and Hirschberg, R. (1989). Changes in gene expression during nitrogen starvation in *Anabaena variabilis* ATCC29413. J. Bacteriol. *171*, 1309-1313.

Webb, C. D., Graumann, P. L., Kahana, J. A., Teleman, A. A., Silver, P. A., and Lostick, R. (1998). Use of time-lapse microscopy to visualize rapid movement of the replication origin region of the chromosome during the cell cycle in *Bacillus subtilis*. Mol. Microbiol. *28*, 883-892.

Weinstock, G. M., McEntee, K., and Lehman, I. R. (1979). ATP-dependent renaturation of DNA catalysed by the *recA* protein of *Escherichia coli*. Proc. Natl. Acad. Sci. (USA) *76*, 126-130.

Williams, J. G. K. (1988). Construction of specific mutations in photosystem II photosynthetic reaction centre by genetic engineering methods in *Synechocystis* 6803. Methods Enzymol. *167*, 766-778.

Williams, J. G. K., and Szalay, A. A. (1983). Stable integration of foreign DNA into the chromosome of the cyanobacterium *Synechococcus R2*. Gene *24*, 37-51.

Williamson, D. H., Gardner, M. J., Preiser, P., Moore, D. J., Rangachari, K., and Wilson, R. J. M. (1994). The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. Mol. Gen. Genet. *243*, 249-252.

Williamson, D. H., Wilson, R. J. M., Bates, P. A., McCready, S., Perler, R., and Qiang, B.-U. (1985). Nuclear and mitochondrial DNA of the primate parasite *Plasmodium knowlesi*. Mol. Biochem. Parasitol. *14*, 199-209.

Wilson, D. B., Garnham, P. C. C., and Swellengrebel, N. H. (1950). A review of hyperendemic malaria. Trop. Dis. Bull.*47*, 667-697.

Wilson, R. J., Williamson, D. H., and Preiser, P. (1994). Malaria and other apicomplexans: the "plant" connection. Infect. Agents Dis. *3*, 29-37.

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

Wilson, R. J. M., Fry, M., Gardner, M. J., Feagin, J. E., and Williamson, D. H. (1992). Subcellular fractionation of the two organelle DNAs of malaria parasites. Curr. Genet. *21*, 405-408.

Wilson, R. J. M., Gardner, M. J., Feagin, J. E., and Williamson, D. H. (1991). Have malaria parasites three genomes? Parasitol Today *7*, 134-136.

Wilson, R. J. M., Gardner, M. J., Rangachari, K., and Williamson, D. H. (1993). Extrachromosomal DNAs in the Apicomplexa. In:*Toxoplasmosis* NATO ASI Series H, J. E. Smith, ed. (Berlin: Springer Verlag), pp. 51-60.

Wilson, R. J. M., and Williamson, D. H. (1997). Extrachromosomal DNA in the Apicomplexa. Microbiol. Mol. Biol. Rev. *61*, 1-16.

Wirth, D. (1998). Malaria: A 21st century solution for an ancient disease. Nature Med. *4*, 1360-1362.

Wittpoth, C., Kroth-Pancic, P. G., and Strotmann, H. (1996). Over-expression and localization of an unknown plastid encoded protein in the diatom *Odontella sinensis* with similarities to a subunit of ABC-transporters. Plant Sci. *114*, 171-179.

Wolfe, K. H., Li, W.-H., and Sharp, P. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. (USA) *84*, 9054-9058.

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

Wolk, C. P., Vonshack, A., Kehoe, P., and Elhai, J. (1984). Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. Proc. Natl. Acad. Sci. (USA) *81*, 1561-1565.

Wu, H. Y., and Kuchka, M. R. (1995). A nuclear supressor overcomes defects in the synthesis of the chloroplast *psbD* gene product caused by mutations in two distinct nuclear genes of *Chlamydomonas*. Curr. Genet. *27*, 263-269.

Wyman, M., Gregory, R. P. F., and Carr, N. G. (1985). Novel role for phycoerythrin in a marine cyanobacterium, *Sycnechococcus* strain DC2. Science *230*, 818-820.

Yagi, T. (1991). Bacterial NADH-quinone oxidoreductases. J. Bioenerg. Biomem. *23*, 211-225.

Yamanaka, K., Ogura, T., Niki, H., and Higara, S. (1996). Identification of two new genes, *mukE* and *mukF*, involved in chromosome partitioning in *Escherichia coli*. Mol. Gen. Genet. *250*, 241-251.

Yamanaka, K., Ogura, T., Niki, H., and Hiraga, S. (1992). Identification and characteriztion of the *smbA* gene, a supressor of the *mukB* null mutant of *Escherichia coli*. J. Bacteriol. *174*, 7517-7526.

Yamazoe, M., Onogi, T., Sunako, Y., Niki, H., Yamanaka, K., Ichimura, T., and Hiraga, S. (1999). Complex formation of MucB, MucE and MukF proteins involved in chromosome partitioning in *Escherichia coli*. EMBO J. *18*, 5873-5884.

Yao, Y., Tamura, T., Wada, K., Matsubara, H., and Kodo, K. (1984). *Spirulina* ferredoxin-NADP+ reductase. The complete amino acid sequence. J. Biochem. (Tokyo) *95*, 1513.

Yates, J. R. (2000). Mass spectrometry: from genomics to proteomics. Trends Genet. *16*, 5-7.

Yu, J., and McIntosh, L. (1998). Isolation and genetic characterization of pseudorevertants from site-directed PSI mutants in *Synechocystis* 6803. Methods Enzymol. *297*, 18-26.

Yu, X. C., and Margolin, W. (1999). FtsZ ring clusters in *min* and partition mutants: role of both the Min system and the nucleoid in regulatinf FtsZ ring localization. Mol. Microbiol. *32*, 315-326.

Zhang, Z., Green, B. R., and Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. Nature *400*, 155-9.

Zhao, G., Meier, T. I., Hoskins, J., and Jaskunas, S. R. (1999). Penicillin-binding protein 2a of *Streptococcus pneumoniae*: expression in *Escherichia coli* and purification and refolding of inclusion bodies into a soluble and enzymatically active enzyme. Protein Exp. Purif. *16*, 331-339.

Zheng, L., White, R. H., Cash, V. L., and Dean, D. R. (1994). Mechanism for the desulfurization of L-cysteine catalysed by the *nifS* gene product. Biochemistry *33*, 4714-4720.

Zhu, L., and Deutscher, M. P. (1987). tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability. EMBO J. *6*, 2473-2477.