Introduction

INTRODUCTION
1. **Introduction**

1.1. Discovery of the apicoplast

1.2. Genome organization

1.2.1. IR-A sector

1.2.2. IR-B sector

1.3. Evolutionary origin

1.3.1. Analysis of *tufA* and *cox2* gene indicates a green algal ancestry

1.3.2. Analysis of small subunit rRNA places the 35kb plDNA closer to euglenoids than rhodophytes

1.3.3. Evidence for a plastid origin outside the green and red algal lineage

1.3.4. The gene content and gene arrangement on the 35kb plDNA molecule more closely resembles those of red algae

1.4. Apicoplast function

1.4.1. Transcription within the apicoplast

1.4.2. Translation of apicoplast ORFs

1.4.3. Protein import

1.4.4. Primary functions

1.5. Rationale
1.1. Discovery of the apicoplast

Malaria has become a major burden to human health in tropical and subtropical areas. Although four members of the genus *Plasmodium* normally infect humans, nearly all deaths are attributable to *Plasmodium falciparum*. The severity of disease caused by *P. falciparum* is primarily due to its ability to modify the surface of infected red blood cells by inserting parasite proteins. Parasitized erythrocytes bind to the host endothelial cells leading to their accumulation in specific organs such as the brain and to the development of cerebral malaria (World Health Organisation, A Global Strategy for Malaria Control, Geneva, 1993). Although considerable efforts have been devoted to the development of a malaria vaccine, no effective vaccine has been developed so far. Moreover, the parasite's resistance to conventional drugs is growing at an alarming rate, making the treatment difficult. New, efficient drugs are thus urgently needed to combat malaria.

*Plasmodium* species belong to the phylum apicomplexa. In apicomplexan parasites including *Toxoplasma*, *Babesia*, *Theileria* and *Eimeria* three types of genetic elements have been identified. These are nuclear, mitochondrial and a 35kb extrachromosomal DNA (Wilson et al., 1994). Circular, extrachromosomal DNAs in *P. lophurae* were first observed by Kilejian (Kilejian, 1975). Wilson and colleagues then characterized equivalent circular chromosomes from *P. knowlesi* and *P. falciparum*. The circle showed a preference for a cruciform configuration owing to the presence of an inverted repeat (Gardner et al., 1988). Sequence analysis revealed that the DNA circle encoded genes that were prokaryotic in nature thus leading to the initial suggestion that it was mitochondrial (Gardner et al., 1988).

The theory of endosymbiosis (Margulis, 1993) states that both plastids and mitochondria are derived from two bacterial cells that took up residence in eukaryotic cells. While the mitochondria is derived from
aerobic bacteria which brought with it the efficient process of oxidative phosphorylation, the plastids originated from endosymbiotic cyanobacterium, a photosynthetic lineage of prokaryotes. Since the malaria parasite is non-photosynthetic it was initially thought that the extrachromosomal DNA was mitochondrial in nature (Wilson et al., 1994, Gardner et al., 1988). However, identification of a second extrachromosomal genome in *Plasmodium* posed a question mark on this hypothesis. This new genome existed as linear arrays of tandemly repeated 6-7kb element, carrying genes for cytochromes and cytochrome oxidases typical of the mitochondrial genome (Feagin, 1992). Additionally, the 6kb linear genome encoded bacterial-type rRNAs which were different from those encoded by the 35kb circle (Feagin et al., 1997).

Sequence analysis revealed that the 35kb element was similar to chloroplast genomes, containing an inverted repeat of ribosomal RNA genes and genes typically found in chloroplasts but not in mitochondria. These included *rpoB/C* (encoding RNA polymerase subunits B, C1 and C2), *tufA* (encoding translation elongation factor Tu) and *clpC* (or hsp93, predicted to be required for protein import into the apicoplast) (Wilson et al., 1996). The 35kb DNA has also been predicted to encode a complete set of tRNAs, ribosomal proteins and several unidentified open reading frames (Wilson et al., 1996). Using *in situ* hybridization Kohler et al. determined whether the 35kb DNA was found within the nucleus, mitochondria or the cytoplasm (Kohler et al., 1997). By hybridizing extracellular *T. gondii* tachyzoites with digoxigenin-labeled DNA probes that covered 10.5kb of the 35kb DNA, they showed that the 35kb Plastid DNA (plDNA) of *T. gondii* localized to a specific region in the cell, adjacent to the apical end of the parasite nucleus. This organelle was named the **apicoplast**. Thin sections through epon-embedded parasites showed that the organelle was enclosed by four bilayer membranes (Kohler et al., 1997). Waller et al. have demonstrated that during schizogony in *P. falciparum*, the
apicoplast acquires a complex branched shape which persists till cytokinesis. Mature schizonts, immediately prior to merozoite release, consist of multiple discreet, rod-shaped apicoplasts. Upon the release of the parasite, each merozoite inherits one apicoplast (Waller et al., 2000).

The apicomplexan plastid contains the smallest plastid genome described so far. It lacks all genes for electron transport complexes typically found in plastids of photosynthetic organisms (Wilson and Williamson, 1997). The presence of multiple surrounding membranes is consistent with a secondary endosymbiotic origin (Wilson et al., 1994, Kohler et al., 1997). Phylogenetic analysis of sequence data for plastid tufA allies apicoplast with plastids of green alga (Kohler et al., 1997). On the other end, organization of ribosomal protein genes is more congruent between apicoplast and non-green plastids (red algae, cryptomonads and chromophytes) (McFadden and Waller, 1997). Dinoflagellates are thought to be the nearest relatives of apicomplexans on the basis of structural similarities (Kohler et al., 1997).

An important question is why the apicoplast has been retained in a highly specialized group of intracellular parasites? The apicoplast genome, although highly homologous to plastid genomes of plants and algae, is highly reduced and completely lacks the genes involved in photosynthesis. The remnant genes can be ascribed to housekeeping functions such as transcription and translation. One possible explanation for the maintenance of the plastid was purely to replicate and transmit itself in a manner similar to genetic elements like transposable elements. However, compounds that specifically inhibited plastid replication reduced parasite viability (Fichera and Roos, 1997) thus indicating that the apicoplast is necessary for parasite survival.
1.2. Genome organization

The 35kb apicoplast genome of *T. gondii* and *P. falciparum* has been completely mapped and sequenced. Apicoplast DNA, unlike other plastid genomes, does not carry genes encoding proteins involved in photosynthesis. However, the remaining genes that are largely devoted to expression are organised in almost exactly the same gene order as their equivalents in plastid genomes. Certain features shared by plant and algal plastid genomes and the 35kb pIDNA of *Toxoplasma* and *Plasmodium* are: circular genome, uniparental inheritance, presence of inverted repeats, polycistronic mode of transcription, A+T richness and a plastid super-operon. The presence of *tufA*, *clpC* (chaperonin), *rpoB/C* and group I intron (*tm1*) (McFadden and Waller, 1997) is also a feature shared with plastid genomes. The 35kb pIDNA has a very high A+T content of about 86%. It contains an inverted repeat of large and small subunit rRNA genes found in chloroplasts but not in mitochondria. Besides, the 35kb pIDNA encodes 25 species of tRNAs, eubacterial RNA polymerases and ribosomal proteins (Fig. 1.1) (Wilson *et al.*, 1996; Wilson and Williamson, 1997).

The inverted repeat (IR) covers about one third of the DNA circle and encodes duplicated large and small subunit (LSU and SSU) rRNA genes (Gardner *et al.*, 1991a), along with nine duplicated tRNA genes (Gardner *et al.*, 1994). Sequence analysis shows that the rRNA genes are not closely linked to those of mitochondria and nucleus (Feagin *et al.*, 1992). The IR extends for 36 nucleotides beyond the *tm1* genes just downstream of the LSU rRNA genes and includes the first three codons of ORF470 (IR-A) and *rps4* (IR-B).
**Fig. 1.1.** Genomic organization of the 35kb plDNA in *Plasmodium falciparum* and *Toxoplasma gondii*. Red indicates features present in *T. gondii* that are absent in *P. falciparum*. The red open circles represent in-frame UGA codons that are predicted to encode tryptophan. Filled circles represent in-frame stop codons (UAA and UAG). Green indicates features present in *P. falciparum* that are absent in *T. gondii* (from the website http://e2kroos.upenn.edu).
1.2.1. IR-A sector

In the IR-A sector, immediately downstream and lying on the same strand as LSU rRNA, lie three putative ORFs. These are ORF470, ORF101 and ORF51. ORF470 corresponds to a highly conserved sequence (ycf24) recorded from the plastids of red alga Porphyra purpurea and Cyanidium caldarium and the diatom Odontella sinensis. At the amino acid level the identity of these sequences with the malarial genes ranges from 47 to 52%. ORF470 has subsequently been shown to correspond to the suJB gene of E. coli (Ellis et al., 2001). At the 3' end of the three ORFs and on the same strand lie rpoB, rpoC1 and rpoC2 which encode for the $\beta$, $\beta'$ and $\beta''$ subunits of RNA polymerase, respectively. These genes are similar to those found in cyanobacteria and chloroplasts and not in mitochondria (Gardner et al., 1991b). The rpo genes provided one of the first clues of the plastid ancestry of the DNA circle (Gardner et al., 1991b). The complete sequence of rpoC shows that it lacks the intron typical of higher plants. Further, rpoC is split into rpoC1 and rpoC2 as in other plastid and cyanobacterial genomes (Wilson et al., 1996). The level of conservation of the predicted peptide encoded by rpoB and rpoC is not as high as ORF470, however, all the known functional domains are conserved in the predicted malarial peptide (Wilson and Williamson, 1997). The rpoA and rpoD coding for the $\alpha$ subunit and the $\sigma$ subunit of the RNA polymerase, respectively are nuclear encoded (McFadden and Roos, 1999). Like other plastid genomes, downstream of the rpoC gene lies the ribosomal protein gene rps2. However, unlike other plastid genomes, atp genes do not follow it. rps2 marks the cross-over point for the direction of transcription from the two arms of the inverted repeat (Wilson et al., 1996) (Fig. 1.1).
1.2.2. IR-B sector

In the IR-B sector, at the 3' of \( tm^T \) is an ORF identified as the ribosomal protein gene \( rps4 \). It shares the first three codons of ORF470 at the other end of the rDNA palindrome. It encodes one of the rRNA binding proteins that initiate the assembly of the 30S ribosome. It has a high A+T content of 94% and only the first 20 amino acids and a large central portion show any similarity to other versions of this protein. Downstream to \( rps4 \) are a cluster of ten tRNA genes. These are \( tm^H, tm^C, tm^L, tm^M, tm^Y, tm^S, tm^K, tm^E \) and \( tm^P \). The leucine tRNA holds the only intron so far identified on the circle. Downstream of the tRNA genes lie a series of ORFs encoding ribosomal proteins, arranged in a manner similar to other plastid genomes. The first ORF in this series is \( rpl4 \) followed by \( rpl23 \) which encodes a poorly conserved peptide. This is followed by \( rpl2 \) which commences with an ATC codon like other plant homologues. The C-terminus of the predicted malarial peptide contains a block of conserved amino acids but is otherwise truncated at both the ends (Wilson et al., 1996). Downstream to it lie \( rps19, rps3, rpl16 \) (32% protein identity with \( E. coli \)) and \( rps17 \) corresponding to the S10 operon. After \( rps17 \) lies \( rpl14 \) that is relatively well conserved (24% homology with \( E. coli \)). Other genes in this sector are \( rps8, rpl6, \) and \( rps5 \). \( rps5 \) is poorly conserved, with only the central region of the predicted peptide showing similarity to other versions (35% identity with \( E. coli \) for this region). After this lies a putative ORF91 followed by \( rpl36 \) encoding a relatively highly conserved peptide (47% identity with \( E. coli \)) despite the open reading frame's marked A+T bias (85%). Downstream of this spc-like operon lies \( rps11 \), a member of the alpha operon of \( E. coli \). After \( rps11 \) lie a pair of ribosomal protein genes, \( rps12 \) and \( rps7 \). \( rps12 \) gene is the best conserved of all the malarial small subunit \( rps \) showing 50% identity with \( E. coli \) (Wilson et al., 1996) (Fig. 1.1).
As in other algal genomes, the ribosomal protein genes in the IR-B sector precede a *tufA* gene, which encodes the elongation factor Tu, a G-protein important for the elongation step of protein synthesis. The predicted peptide is highly divergent sharing only 45% amino acid identity with the *tuf* genes of *E. coli* and 51% identity with *Anacystis nidulans* and *Euglena gracilis*. However, several highly conserved functional domains are evident, including the four clusters of residues in domain I involved in GTP binding. The residues defining the GDP binding pocket are also conserved. Despite the high A+T content of *tufA*, it encodes one of the best conserved proteins on the circle. In a less well conserved region topologically close to the GTP binding domain, the malarial sequence has a specific insertion like other plastid versions of EF-Tu (Wilson *et al.*, 1996).

Downstream of *tufA* lie four tRNA genes. Another short ORF, ORF129 then leads to the final ORF on the IR-B single copy region. This has been identified as *clpC*, a member of the *hsp100* family (now annotated as hsp93). The gene is believed to code for a molecular chaperone that aids in the import of nuclear-encoded proteins targeted to the apicoplast lumen (Foth *et al.*, 2003). It corresponds by sequence similarity to the double nt-binding, regulatory forms of *clp* rather than the single nt-binding subfamilies *clpX* and *Y* (Gottesman *et al.*, 1993). However, only the second of the two ATP-binding domains is conserved in the predicted malarial peptide. Alignments of amino acids from double nt-binding subunits of clp proteins showed little similarity with the malarial sequence throughout the first ATP-binding domain. In contrast, a high level of similarity was evident in the second nt-binding domain. Following the *clpC* gene are present two tRNA genes. They are separated by 240 nucleotides that contain an unassigned ORF, OFR79. Then lies the ORF105 that overlaps the *rps2* gene on the opposite strand (Wilson *et al.*, 1996).
Apicomplexan plDNA (from *Plasmodium*, *Toxoplasma* and *Eimeria*) is conserved at the levels of gene content, gene order, and intergenic sequences, supporting the contention that they have evolved from a single source.

1.3. **Evolutionary origin**

The apicoplast is present in all the three major apicomplexan lineages: haemosporins (*Plasmodium*), Piroplasms (*Babesia* and *Theileria*) and coccidians (*Toxoplasma*, *Eimeria*, *Hepatozoon* and *Sarcocystis*) (McFadden et al., 1997). The only apicomplexans believed to lack an apicoplast are *Colpodella* and *Cryptosporidium parvum* (Foth and McFadden, 2003). It has been suggested that these lineages have diverged from their last common ancestor, which possessed a plastid several hundred million years ago. It has been very hard to trace the evolutionary origin of the apicoplast. Phylogenetic analysis of the 35kb plDNA of *P. falciparum* has proven difficult because of long distances between its DNA sequence and those of other organisms. It is suggested that these distances are due to the high A+T content of the apicoplast genome. Plastids are usually categorized by pigmentation, which is lacking in *Plasmodium* and *Toxoplasma*.

Another important character for determining evolutionary history is the number of membranes bounding the plastids. Thus, plastids with two membranes, such as those of red algae, green algae, plants and glaucophytes are thought to derive from a single primary endosymbiosis of a cyanobacterium. On the other hand, plastids with more than two bounding membranes such as diatoms, dinoflagellates, euglenoids and cryptomonads are probably derived from secondary endosymbiosis in which phagotrophic eukaryotes engulfed and retained photosynthetic eukaryotes (McFadden and Waller, 1997). The sharpest images of the *Toxoplasma* plastid show four surrounding membranes (Kohler et al., 1997) indicating the secondary endosymbiotic origin of the apicoplast. On the basis of gene content
and gene structure the apicomplexans are related to the red lineage with many of the ribosomal protein genes forming a super operon as in the red lineage. On the basis of structural similarities and phylogenetic analysis of nuclear genes, apicoplast is closely related to dinoflagellates (Kohler et al., 1997). Dinoflagellates are a diverse and abundant group of marine or aquatic unicellular protozoa of very ancient origin. They enter into symbiotic associations with a wide range of invertebrates. However, in some cases dinoflagellates assume a parasitic life cycle by taking advantage of the host. Modern apicomplexans also parasitize a wide range of invertebrates. Hence, it is possible that the early apicomplexans shared the dinoflagellate's ability to interact mutualistically with invertebrates and probably some abandoned photosynthesis in preference to parasitism (McFadden and Waller, 1997).

1.3.1. Analysis of *tufA* and *cox2* gene indicates a green algal ancestry

Analysis of the *tufA* gene sequence from *P. falciparum*, *T. gondii* and *E. tenella* places the apicomplexan 35kb element solidly within the plastid. The similarity of apicomplexan and plastid *tufA* genes is also supported by the presence of two insertions characteristic of plastids and cyanobacteria, although the length of these insertions is variable among the apicomplexa (Kohler et al., 1997). The *tufA* gene sequence shows significant amino acid identity with *tuf* genes of *E. coli* and *Euglena gracilis*. Several highly conserved functional domains are evident, including the four clusters of residues present in domain I involved in GTP binding. Although the A+T content of *tufA* is very high, yet it encodes one of the best conserved proteins specified by the circle (Wilson et al., 1996).

Recent studies by Funes et al. (Funes et al., 2002) have suggested a green algal ancestry based on analysis of the *cox2* gene, which encodes COXII, a subunit of mitochondrial cytochrome c oxidase. In
apicomplexans the COXII is nuclear-encoded (Gardner et al., 2002). However, in other organisms, with the exception of certain green algae and leguminous plants, it is encoded by the mitochondrial genome (Gray, 1999; Palmer et al., 2000). The COXII protein of apicomplexan parasites contains two polypeptides which correspond to the amino terminal and the carboxyl terminal domains of the canonical COXII, the two domains being encoded by two nuclear genes, cox2a and cox2b (Funes et al., 2002). This gene separation is also found in certain green algae where it appears that the cox2 gene split in the mitochondrial DNA before cox2a and cox2b were transferred to the nucleus (Funes et al., 2002). Funes et al. presented a phylogeny of COXII indicating that the apicomplexan genes are most closely related to cox2 genes of green algae. They also suggest that apicomplexans acquired their split cox2a and cox2b genes through lateral gene transfer, nucleus to nucleus, from the endosymbiotic green alga that gave rise to the plastid.

1.3.2. Analysis of small subunit rRNA places the 35kb plDNA closer to euglenoids than rhodophytes

Phylogenetic analysis of a portion of 35kb plastid small sub-unit rDNA has suggested that it is more closely related to euglenoid plastids rather than rhodophytes (Egea and Lang Unnasch, 1995). The T. gondii organellar SSU rDNA was initially aligned with SSU rDNA genes from bacteria, plastids and two other apicomplexan parasites. Log-det transformation analysis showed that the T. gondii, P. falciparum and B. bovis SSU rDNA formed a monophyletic cluster. Significantly, the plastid SSU rDNA of the euglenoids, Euglena and Astasia, appeared near the base of the apicomplexan branch of the tree. There was no indication that the rhodophyte plastids, Antithamnion and Cyanidium, are more closely related to apicomplexans than are other algal plastids, such as those of phaeophytes (Pylaiella) or chrysophytes (Olisthodiscus) (Fig. 1.2). For further analyses, a subset of the SSU rDNA sequences was chosen.
Again, the euglenoid plastids and the apicomplexan organelles formed sister groups. Several other methods of phylogenetic analysis including maximum likelihood, distance matrix and parsimony methods indicated the same clustering of apicomplexan sequences with those of the euglenoids rather than the rhodophytes (Egea and Lang Unnasch, 1995).

1.3.3. Evidence for a plastid origin outside the green and red algal lineage

An analysis by Blanchard and Hicks (Blanchard and Hicks, 1999) used apicoplast genomic characters to trace the evolution of apicomplexa. Using both primary sequence characters (nucleotides) and genomic characters (gene content, intron presence and genomic structure) present in all completely sequenced plastid genomes, they attempted to provide a stable phylogenetic position to apicomplexa. Their analysis revealed that apart from the presence of a super operon as in the red lineage, the \textit{rps2-rpoB-rpoC1-rpoC2} gene order is also conserved among \textit{Cyanophora}, \textit{Porphyra}, \textit{Odontella}, \textit{Plasmodium} and plants. \textit{Chlorella} and \textit{Euglena} have a different gene order. Blanchard and Hicks also conducted a cladistic analysis of gene content in the ribosomal gene clusters using the presence of the gene in \textit{Synechocystis} as the ancestral state. The analysis placed the \textit{Plasmodium} as a sister group to green algae and plants and is supported by the loss of \textit{rpl14}, \textit{rps17}, \textit{rpl16}, \textit{rps5} in all green algal and plant lineages. \textit{Plasmodium} contains two other genes, \textit{clpC} and \textit{ycf24}, that are found in \textit{Porphyra}, \textit{Odontella} and \textit{Cyanophora}, but not in \textit{Chlorella}, \textit{Euglena} and plants (Blanchard and Hicks, 1999).
Introduction

The grouping of *Plasmodium* with *Odontella* and *Porphyra* is based on gene content and gene structure of the apicoplast DNA (modified from Blanchard and Hicks, 1999).

Phylogenetic analysis of plastid SSU rDNA suggests that it is more closely related to euglenoid plastids rather than rhodophytes (modified from Egea and Lang Unnasch, 1995).

**Fig. 1.2.** Evolutionary origin of the 35kb plDNA
Plants have numerous plastid introns and there are over a hundred independently derived introns in *Euglena*. However, there are only two introns in *Chlorella* and a single plastid intron in *Cyanophora* and *Plasmodium*, a group I intron in a tRNA-leu gene, that was probably present in the cyanobacterial endosymbiont (Blanchard and Hicks, 1999). This analysis supports a plastid origin outside the green and red lineages. Blanchard and Hicks favour a position in which *Plasmodium* is placed close to *Odontella* (a brown alga) because of the likelihood of chlorophyll-c containing dinoflagellate plastids and *Odontella* plastids sharing a common ancestry based on their light-harvesting proteins (Fig. 1.2). The study suggests that a non-green source cannot be excluded.

1.3.4. The gene content and gene arrangement on the 35kb plDNA molecule more closely resembles those of red algae

Many ribosomal protein genes form a super operon in the red lineage. Based on the arrangement of genes in the super operon, apicoplast DNA is closer to the red algal lineage (McFadden and Waller, 1997). In plastids and cyanobacteria the four operons (S10, str, spc and alpha) are amalgamated into two or even one super operon. In *Synechocystis*, *Chlorella* and plants these ribosomal proteins are found in two separate clusters from *rpl3* to *rpl31*. In *Cyanophora* the *rpl3* to *rpl31* cluster is split between *secY* and *rpl36* resulting in three clusters. In comparison, all plastids in the red lineage and *Plasmodium* have a single contiguous cluster (Blanchard and Hicks, 1999). The plastid genomes of red algae, cryptomonads and diatoms have the str operon transposed to the rear of other operons. The malarial plastid genome shows a significant similarity to red algae, cryptomonads and diatoms in this respect (McFadden and Waller, 1997).

Several nuclear-encoded cytosolic proteins with plant like sequences such as Glucose-6-phosphate isomerase and enolase have been described for *Toxoplasma* and *Plasmodium*. It has been suggested that
resolution of plastid's origin might not come from plastid genes but from those translocated to the nucleus. The latter would be under tighter evolutionary constraint than in the organelles, where the rate of genetic drift can be increased for several reasons (Blanchard and Lynch, 2000). An analysis of the nuclear-encoded and apicoplast targeted Glyceraldehyde-3-phosphate dehydrogenase was carried out by Fast et al. (2001). Sequencing of cytosolic and plastid copies of GAPDH from *T. gondii*, several ciliates and the heterokont alga *Heterosigma akashiwo* was carried out. The analysis was based on the information that the plastid-targeted GAPDH gene of dinoflagellates is related to that of cryptomonads. However, in both cases the GAPDH gene is a replacement copy of itself and is a duplicate version, not a cyanobacterial version like those in plants or algae. Construction of phylogenetic trees with apicomplexans, ciliates and heterokont GAPDH sequences showed that the plastid-targeted *Toxoplasma* GAPDH is most closely related to the cytosolic GAPDH found in the dinoflagellate. This suggests that rather than being independent, the plastids of *Toxoplasma* and dinoflagellates originate from a common endosymbiotic event involving a red alga.

The hypothesis that the apicomplexan plastid has a red algal origin is now better accepted. This can be further confirmed by phylogenetic analysis of other nuclear-encoded genes whose products are targeted to the apicoplast.

1.4. Apicoplast function

Although the apicoplast genome is presumed to be the remnant of a much larger precursor, certain features of the genome point to its functional role in the parasite. Apicoplast ORFs have been maintained despite extensive sequence divergence and the genetic content of the circle has been conserved across various genera of the apicomplexa (Wilson and Williamson, 1997).
1.4.1. Transcription within the apicoplast

More direct evidence of apicoplast genome functionality has been provided by the transcriptional activity of the organelle. RNase protection assays, RT-PCR and northern blot analysis methods have identified transcripts for tRNAs, rpoB, rpoC1 and rpoC2, LSU and SSU rRNA, tufA, clpC and ORF470 (Wilson et al., 1996; Feagin and Drew, 1995; Gardner et al., 1991a; Gardner et al., 1991b; Gardner et al., 1993; Preiser et al., 1995). The fact that these genes are transcribed strengthens the view that apicoplast is a functional organelle.

1.4.2. Translation of apicoplast ORFs

There is evidence to suggest that apicoplast has an active protein synthesis machinery. Nearly all the genes present on the 35kb circle specify components required for protein synthesis. Although plastid contents are largely homogeneous, particulate structures comparable to the size of 70S ribosomes of plastids, mitochondria and bacteria are present within the organelle (McFadden et al., 1997 and McFadden et al., 1996). The detection of polysomes by hybridization to the plastid rRNAs and mRNA (Roy et al., 1999) is also consistent with the plastid genome encoding components of ribosomes and supports the idea that protein synthesis is active in the apicoplast.

Further indirect evidence has been provided by the use of antibiotics. Prokaryotic translation inhibitors such as thiostrepton, clindamycin, azithromycin and chloramphenicol (McFadden and Roos, 1999) have been shown to inhibit the parasite growth. In vitro, thiostrepton binds preferentially to the GTPase domain of plastid 23S rRNA rather than to that of cytosolic rRNAs (Clough et al., 1997; McConkey et al., 1997). Protein synthesis within the apicoplast is believed to be the target of these drugs although this remains to be confirmed.

Recent work suggests that plastid protein synthesis is important for housekeeping functions. There are two large conserved ORFs on
plastid DNA. These encode ORF470 and clpC in *P. falciparum*. The chaperone clpC, a class I clp/Hsp100 ATPase [now classified as Hsp93 (Jackson-Constan *et al.*, 2001)] is found universally in plastids as part of the machinery for processing imported peptides (Nielson *et al.*, 1997). It is assumed to play the same role in apicomplexa. ORF470 is an orthologue of the hypothetical chloroplast frame, *ycf24*. Homology studies indicate that it corresponds to the *sufB* gene of *E. coli*. The bacterial *suf* operon comprises six genes (*sufA, sufB, sufC, sufD, sufS* and *sufE*) in its complete form. Knock-out experiments carried out in *E. coli* implicated several of these genes in iron homeostasis/assembly of [Fe-S] clusters and resistance to oxidative stress (Patzer and Hantke, 1999 and Nachini *et al.*, 2001). A candidate of *sufC* has been found on chromosome 14 of *P. falciparum*. It has a putative plastid-targeting leader sequence. This suggests the possibility that products of *sufB* and *sufC* might interact in apicomplexan plastid (Wilson, 2002). It is proposed that *sufB* and *sufC* are required for the assembly of [2Fe-2S] clusters in the plastid organelle to convert imported apopferredoxin to the holoprotein (Wilson, 2002).

These observations suggest the presence of an active but minimal protein synthesis system in apicomplexan plastids.

### 1.4.3. Protein import

Nuclear genes whose products are targeted to the apicoplast have been identified from *P. falciparum* and *T. gondii* (Waller *et al.*, 1998). Ribosomal protein genes like *rps9* and *rpl28* that are missing from the plastid genome have been located on chromosomal DNA. These genes carry an N-terminal bipartite sequence that targets their products to the plastid (Waller *et al.*, 1998 and Yung *et al.*, 2001). It is estimated that between 1000 to 5000 proteins in plant chloroplasts are encoded by nuclear DNA (Martin and Hermann, 1998) and, by analogy, most of the protein content of the apicomplexan apicoplast is likely to be nuclear-encoded.


**Introduction**

*T. gondii* and *P. falciparum* apicoplasts have three and four surrounding membranes, respectively (Kohler et al., 1997; Foth and McFadden, 2003). In the case of triple-membraned plastids, the outermost membrane is likely to be derived from the host endomembrane system because signal peptides and the endomembrane pathway are used for trafficking peptides to the organelle (Sulli and Schwartzbach, 1995). Organisms having secondarily acquired plastids that are surrounded by additional membranes direct products to the organelle via the secretory pathway. In such cases, an N-terminal signal peptide precedes the transit peptide (Lang et al., 1998). Apicomplexans seem to have a similar sorting machinery (Fig. 1.3).

Studies by Waller et al. (1998) have led to the identification of certain nuclear-encoded gene products that are targeted to the apicoplast. These include ribosomal proteins S9 and L28 as well as proteins involved in type II fatty acid biosynthesis e.g. Fab Z (β-hydroxyacyl ACP dehydratase), Fab H (β-ketoacyl ACP synthase) and ACP (acyl carrier protein). Fab I (enoyl-ACP reductase) has also been localized in the apicoplast (Surolia and Surolia, 2001). In addition, enzymes of the non-mevalonate pathway of isoprenoid biosynthesis, DOXP synthase and DOXP reductoisomerase, are also targeted to the apicoplast (Jomaa et al., 1999). The N-terminal presequence has been used to search the *P. falciparum* genome sequence and about 466 proteins have now been predicted to be apicoplast targeted (Foth et al., 2003).

The N-terminal bipartite pre-sequence of proteins that are targeted to the apicoplast carries two domains: a signal peptide and a transit peptide (Waller et al., 1998). The signal peptide consists of a short hydrophobic domain followed by a Von-Heijne-type motif. Following the signal peptide is the transit peptide which, like its plant counterparts, carries a net positive charge (Waller et al., 1998; Foth et
al., 2003). Unlike plants, the complement of hydroxylated serine and threonine residues is low in the malarial transit peptides, whereas those of *Toxoplasma* are more typical. In *Plasmodium*, the transit peptide is enriched in lysine and asparagine residues (Foth et al., 2003).

Gene fusion experiments carried out by Waller et al. have suggested that protein targeting to the apicoplast is an ordered two-step process. Proteins must first enter the secretory pathway. Once there, the transit peptide would mediate final transfer into the lumen of the apicoplast. Studies indicate that components of the bipartite leader are removed in a two-step sequential fashion (Waller et al., 2000). The signal peptide is removed first followed by the transit peptide. Another factor suggested to play a role in protein targeting to plant plastids is the binding of Hsp70 chaperones to plastid transit peptides (Ivey and Bruce, 2000; Ivey et al., 2000), although Hsp70 binding to apicoplast transit peptides remains to be demonstrated (Matambo et al., 2004). An orthologue of the plant gene encoding the plastid stromal processing peptidase (SPP) that removes transit peptides from imported proteins, is present on chromosome 14 of *P. falciparum* and *P. yoelii* (Van Dooren et al., 2002).

On the basis of studies carried out so far, a model for protein import into the apicoplast has been proposed (Foth et al., 2003). In this model, the signal peptide mediates co-translational insertion into the endomembrane lumen and is cleaved. Endomembrane derived vesicles then dock with the outermost membrane of the apicoplast, delivering the proteins whose NH₂-termini are maintained in an unfolded conformation by bound Hsp70 molecules.
Fig. 1.3. Protein targeting to the apicoplast (redrawn from Waller et al., 1998). (A) Uptake of prokaryote by a eukaryote (primary endosymbiosis). (B) Targeting of nuclear-encoded gene products to the primary endosymbiont requires an N-terminal transit peptide. (C) The heterotrophic eukaryote phagocytosing a photosynthetic eukaryote produced by primary endosymbiosis (secondary endosymbiosis). (D) Targeting of nucleus-encoded gene products to secondary plastid requires an N-terminal signal peptide followed by a transit peptide. N', nucleus of eukaryote phagocytosing the prokaryote; N'', nucleus of heterotrophic eukaryote; P, plastid; S, signal peptide; T, transit peptide.
The positively charged transit peptides are drawn through negatively charged transmembrane pores into the reducing environment of the apicoplast lumen. Apicoplast-encoded ClpC (Hsp100/Hsp93) then binds to the transit peptide preventing retrograde movement and drawing the protein into the apicoplast. The transit peptide is then cleaved by a stromal peptidase (Van Dooren et al., 2002) and the mature protein refolds with the assistance of the apicoplast targeted GroEL homolog Cpn60 (Gardner et al., 2002).

1.4.4. Primary functions
Apicoplast is believed to be the site for type II fatty acid biosynthesis, isoprenoid biosynthesis as well as synthesis of heme within the parasite (Waller et al., 1998; Jomaa et al., 1999; Dhanasekaran et al., 2004).

There are two types of fatty acid biosynthesis, type I and type II. Type I is found in the cytosol of animals and fungi. Type II is found in bacteria and is restricted to the plastids of plants and algae in eukaryotes. Mammalian enzymes (type I) are part of a multi-domain polypeptide which includes ACP, acetyl-CoA-ACP transacylase, \( \beta \)-ketoacyl-ACP synthase, \( \beta \)-hydroxyacyl-ACP dehydratase and enoyl-ACP reductase. In contrast, the cytosolic bacterial enzymes and the plastidic enzymes in plants (type II) are discreet mono-functional proteins. In Toxoplasma and Plasmodium, nuclear-encoded genes that resemble genes encoding proteins involved in type II fatty acid biosynthesis have been identified. These include ACP, FabZ and FabH (Waller et al., 1998). When fused with GFP, these proteins were found to localize in the apicoplast (Waller et al., 1998). All these proteins bear N-terminal pre-sequences consistent with apicoplast targeting and are members of the fatty acid synthase multi-enzyme complex.
Differences between type I and type II enzymes are the basis for the selectivity of a number of antibacterials including thiolactomycin and triclosan. The antibiotic thiolactomycin is a selective inhibitor of type II fatty acid biosynthesis. In *E. coli* it inhibits the condensing enzymes Fab B, Fab F and Fab H (Nishida *et al.*, 1986) and also inhibits the equivalent plastid enzymes in plants. In contrast, thiolactomycin has no effect on type I fatty acid biosynthesis (Waller *et al.*, 1998). Thiolactomycin inhibits growth of *in vitro* cultures of *P. falciparum* with an IC$_{50}$ of about 50µM (Waller *et al.*, 1998). This level of inhibition is comparable with that seen in isolated pea and spinach plastids. Thiolactomycin inhibition of malaria growth thus provides additional support that apicoplast is the site for type II fatty acid biosynthesis.

Surolia and Surolia (2001) have demonstrated the antimalarial effect of triclosan on *P. falciparum*. Triclosan is a selective inhibitor of bacterial Fab I (Enoyl ACP reductase) and inhibits *P. falciparum* growth *in vitro*. Efficacy of triclosan has also been examined *in vivo* in *P. berghei* and a single subcutaneous injection of 38mg/kg completely clears the parasite from circulation. The Fab I of *P. falciparum* has been purified and characterized. Additionally, triclosan has been shown to bind to it and inhibit its activity (Surolia and Surolia, 2001). These studies point strongly towards the involvement of apicoplast in type II fatty acid biosynthesis.

Two proteins of the alternative non-mevalonate pathway of isoprenoid biosynthesis, DOXP synthase and DOXP reductoisomerase, have been localized in the apicoplast of *P. falciparum* (Jomaa *et al.*, 1999). The biosynthesis of isoprenoids such as sterols and ubiquinones depends on the condensation of different members of isopentenyl-diphosphate units. In mammals and fungi, isopentenyl-diphosphate is derived from the mevalonate pathway. This pathway depends on the condensation
of three molecules of acetyl-CoA into HMG-CoA, which is reduced to mevalonate by HMG-CoA reductase. Mevalonate is further converted to isopentenyl- diphosphate with mevalonate–5-diphosphate as an intermediate. Previous studies revealed very low HMG-CoA reductase activity in *P. falciparum* (Vial *et al.*, 1984) and attempts to establish HMG-CoA reductase inhibitors as antimalarial drugs were unsuccessful (Crellier *et al.*, 1994).

In higher plants, plastidic isoprenoids such as carotenoids are formed by DOXP pathway or the non-mevalonate type pathway. The DOXP pathway is characterized by the condensation of glyceraldehyde-3-phosphate and pyruvate into DOXP (1-deoxy-D-xylulose-5-phosphate) and its conversion to 2-C-methyl-D-erythritol-4-phosphate by the enzymes DOXP synthase and DOXP reductoisomerase. The gene encoding DOXP reductoisomerase has been identified on chromosome 14 of *P. falciparum*. This gene exhibits significant similarity with known bacterial and blue algal protein sequences (Jomaa *et al.*, 1999). A gene very similar to DOXP synthase has been also identified in *P. falciparum*. Transfection of *T. gondii* with a construct containing the NH₂-terminal of DOXP reductoisomerase fused to GFP led to the localization of protein in apicoplast (Jomaa *et al.*, 1999). The presence of DOXP synthase and DOXP reductoisomerase in the apicoplast suggested its involvement in the non-mevalonate pathway of isoprenoid biosynthesis (Jomaa *et al.*, 1999). Moreover, inhibitors of the DOXP pathway, fosmidomycin and FR 9000098 inhibited the growth of *P. falciparum* in submicromolar concentrations. Additionally, mice infected with *P. vincei* were cured by intraperitoneal injections (10 mg/kg of fosmidomycin or 5 mg/kg of FR-900098) of the two drugs (Jomaa *et al.*, 1999).

Apicoplast has also been implicated in the heme biosynthesis. Recent studies by Dhanasekaran *et al.* (2004) and Varadharajan *et al.* (2004)
have localized the *P. falciparum* delta-aminolevulinate dehydratase (ALAD) and ferrochelatase, the second and last enzymes respectively of the heme biosynthetic pathway to the apicoplast of malaria parasite. Earlier studies by Surolia and Padmanaban (1992) have shown the import of host ALAD from the red cell cytoplasm by intraerythrocytic malaria parasite. Dhanasekaran *et al.* (2004) propose that the *P. falciparum* ALAD may account for 10% of the total ALAD activity, the rest being accounted for by the host enzyme imported by the parasite. Thus, *P. falciparum* ALAD though involved in heme biosynthesis may not account for the total de novo heme biosynthesis in the parasite.

The evidence that apicoplast is involved in the essential functions of type II fatty acid biosynthesis, isoprenoid biosynthesis and heme biosynthesis provides the basis for its persistence in apicomplexan parasites. It also identifies novel drug targets for chemotherapy and strengthens the position of the apicoplast as a putative drug target for malaria.

### 1.5. Rationale

Studies carried out in the last several years have demonstrated a direct link between apicoplast function and intracellular survival of the parasite. A number of inhibitors of prokaryotic transcription and translation have been found to be effective against *Toxoplasma* and *Plasmodium*. The apicoplast genome encodes an RNA polymerase that is homologous to that of cyanobacteria and other eubacteria (Gray and Lang, 1998). While the β, β' and β” subunits of RNA polymerase are apicoplast-encoded, the α subunit and σ subunit are encoded by the nucleus (McFadden and Roos, 1999). The β, β’ and β” polymerase of bacteria and plastids is highly sensitive to rifampicin (Pukrittayakamee *et al.*, 1994) and the antimalarial activity of rifampicin suggests that this drug might block apicoplast
transcription (Gardner et al., 1991b). Several prokaryotic translation blockers also inhibit *P. falciparum* and *T. gondii* growth (Fichera et al., 1995). Lincosamides e.g. clindamycin and macrolides e.g. azithromycin block protein synthesis by interacting with the peptidyltransferase domain of bacterial 23S rRNA. These antibiotics have been shown to inhibit the growth of *P. falciparum* and *T. gondii* (Jeffries and Johnson, 1996). Two thiopeptide antibacterial agents, thioestrepton and micrococcin are potent inhibitors of *P. falciparum* growth *in vitro* (McConkey et al., 1997; Rogers et al., 1998). In *P. falciparum*, only the apicoplast LSU rRNA is susceptible to thioestrepton and it is unlikely to affect cytosolic or mitochondrial rRNAs.

Studies carried out by Fichera and Roos (1997) have demonstrated that replication of the apicoplast genome in *T. gondii* is specifically inhibited by ciprofloxacin (a fluoroquinolone), which is a specific and selective inhibitor of prokaryotic gyrases (Furet and Pechere, 1991). This in turn reduces parasite viability in culture. Ciprofloxacin does not inhibits eukaryotic gyrases or mitochondrial DNA replication. Studies in plant systems also indicate that plastid gyrases are sensitive to fluoroquinolones. Experiments carried out by Fichera and Roos demonstrated that treatment of *T. gondii* intracellular tachyzoites with ciprofloxacin results in specific depletion of extranuclear DNA. This extranuclear DNA co-localized with apicoplast specific DNA probes used in *in situ* hybridization (Kohler et al., 1997). Quantitative hybridization experiments revealed that treatment with 25μM ciprofloxacin, reduced the plastid genome copy number by more than ten-fold over the course of replication within the infected cell. The copy number was further reduced in the second infectious cycle after the parasites lysed out of the initial host cell, producing the 'delayed death phenotype' that was earlier reported for clindamycin and other mechanistically similar drugs (Fichera et al., 1995). Upon treatment
with ciprofloxacin as well, parasite division was inhibited only after entry into the second host cell.

Although inhibition of apicoplast DNA replication has a significant effect on parasite survival, the mechanism of apicoplast replication is not clearly understood. Hence, one of the main objectives of this study was to identify replication initiation sites of apicoplast DNA as an initial step in the analysis of the mechanism of DNA replication. The demarcation of DNA replication initiation sites would further help in the identification of DNA-protein interactions at these sites.

Apart from rRNA, tRNA and ribosomal proteins, several ORFs of unknown function are present in the apicoplast genome. These include tufA, clpC and ORF470. The latter is a homolog of ycf24 and corresponds to the sufB gene of E. coli. Transcriptional regulation of these genes is not understood. Hence, another objective was to carry out the transcriptional analysis of ORF470 to determine its monocistronic/polycistronic nature. This would provide leads toward understanding transcriptional control in the apicoplast. Together, these studies would help gain insight into the biology of the apicoplast. With the above rationale, the primary objectives of the study were identification of the replication origins of the apicoplast genome. This would involve generation of DNA fragments covering the entire apicoplast genome as well as purification of the 35kb apicoplast genome of P. falciparum and isolation of apicoplast genome replication intermediates. Another objective was the transcriptional analysis of ORF470 as a step towards understanding transcriptional regulation in the apicoplast.