Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*

( apicomplexan parasites / organellar targeting / fatty acid biosynthesis / malaria)

ROSS F. WALLER*, PATRICK J. KEELING*, ROBERT G. K. DONALD†, BORIS STRIEPEN‡, EMANUELA HANDMAN§, NAOMI LANG-UNNASCH¶, ALAN F. COWMAN¶, GURDIAL S. BESRA¶, DAVID S. ROOS†, AND GEOFFREY I. MCFADDEN*∥

*Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville VIC 3052, Australia; †Department of Biology, University of Pennsylvania, Philadelphia, PA 19104-6018; ‡Walter and Eliza Hall Institute of Medical Research, Parkville VIC 3050, Australia; §Division of Geographic Medicine, Department of Medicine, University of Alabama, Birmingham, AL 35294-2170; and ¶School of Microbiological, Immunological and Virological Sciences, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, United Kingdom

Communicated by Adrienne Clarke, University of Melbourne, Parkville, Australia, July 27, 1998 (received for review June 12, 1998)

ABSTRACT A vestigial, nonphotosynthetic plastid has been identified recently in protozoan parasites of the phylum Apicomplexa. The apicomplexan plastid, or “apicoplast,” is indispensable, but the complete sequence of both the *Plasmodium falciparum* and *Toxoplasma gondii* apicoplast genomes has offered no clue as to what essential metabolic function(s) this organelle might perform in parasites. To investigate possible functions of the apicoplast, we sought to identify nuclear-encoded genes whose products are targeted to the apicoplast in *Plasmodium* and *Toxoplasma*. We describe here nuclear genes encoding ribosomal proteins S9 and L28 and the fatty acid biosynthetic enzymes acyl carrier protein (ACP), β-ketoacyl-ACP synthase III (FabH), and β-hydroxyacyl-ACP dehydratase (FabZ). These genes show high similarity to plastid homologues, and immunolocalization of S9 and ACP verifies that the proteins accumulate in the plastid. All the putatively apicoplast-targeted proteins bear N-terminal presequences consistent with plastid targeting, and the ACP presequence is shown to be sufficient to target a recombinant green fluorescent protein reporter to the apicoplast in transgenic *T. gondii*. Localization of ACP, and very probably FabH and FabZ, in the apicoplast implicates fatty acid biosynthesis as a likely function of the apicoplast. Moreover, inhibition of *P. falciparum* growth by thiolactomycin, an inhibitor of FabH, indicates a vital role for apicoplast fatty acid biosynthesis.

The fatty acid biosynthesis genes identified here are of bacterial type, and distinct from those of the equivalent pathway in animals, fatty acid biosynthesis is potentially an excellent target for therapeutics directed against malaria, toxoplasmosis, and other apicomplexan-mediated diseases.

The origin of the apicoplast is uncertain (10). Thus far, all apicoplasts examined show common characters consistent with a shared origin (4, 11). The multiple surrounding membranes are consistent with a secondary endosymbiotic origin (3, 12, 13), and phylogenetic analysis of sequence data for the plastid protein TufA weakly allies apicoplasts with plastids of green algae and plants, suggesting a green algal endosymbiont (3). On the other hand, the organization of ribosomal protein genes is more congruent between apicoplasts and nongreen plastids (red algae, cryptomonads, and chromophytes), possibly suggesting a different endosymbiont (7). Dinoflagellates are thought to be the nearest relatives of the apicomplexa (6) but very little is known about their plastids, which may derive from either a primary (14) or secondary endosymbiosis (12).

A key question is why the apicoplast has been retained in a highly specialized group of intracellular parasites. The apicoplast genome, although clearly homologous to plastid genomes of plants and algae, is highly reduced and completely lacking in genes known to be involved in photosynthesis (1, 3, 7, 10, 15). Of the 68 genes in the malaria apicoplast, at least 60 can be ascribed to “house-keeping” functions such as transcription and translation (1, 15). One or more of the remaining eight unidentified ORFs may be involved in some function key to plastid retention. However, plastids also are dependent on many hundreds of nuclear-encoded proteins that are targeted to the organelle (16). To date, no nuclear-encoded plastid proteins have been reported in the apicomplexa despite the absence of numerous essential house-keeping genes from the apicoplast genome (1, 7). We have therefore taken advantage of sequence information from the *Toxoplasma gondii* expressed sequence tag (EST) databases, and the *Plasmodium falciparum* genomic sequence data, to identify candidate plastid proteins encoded by genes in the parasite nucleus. We show here that several of these proteins are indeed targeted to the apicoplast, and their identity provides the first clues to the possible function of this intriguing organelle.

**MATERIALS AND METHODS**

Cloning of cDNAs and gDNAs from *T. gondii* and *P. falciparum*. Publicly available *Toxoplasma* and *Plasmodium* databases were searched for likely plastid-targeted gene sequences. *T. gondii* EST sequences (*Toxoplasma* Genome Web; http://www.ebi.ac.uk/parasites/toxo/toxpage.html) with sim-

Abbreviations: ACP, acyl carrier protein; FabH, β-ketoacyl-ACP synthase III; FabZ, and β-hydroxyacyl-ACP dehydratase; GFP, green fluorescent protein.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF038922–AF038929 and AF067150).

To whom reprint requests should be addressed. e-mail: g.mcfadden@botany.unimelb.edu.au.
Phylogenetic Analysis. Neighbor-joining trees and neighbor-joining bootstrap trees were constructed from distances corrected by the Dayhoff PAM substitution matrix by using the programs NEIGHBOR and PROTDIST (19). Protein maximum likelihood was performed by using PROML (20). Trees were searched exhaustively with the JTT-f substitution model, using constraints derived from nodes that found strong support from the neighbor-joining trees, or otherwise well-defined relationships. These constraints are indicated by asterisks in Fig. 1, with the exception that mycobacterial rps9 genes also were constrained. Bootstrap values for protein maximum likelihood were estimated by using the RELL method, and collated by using MOLCON. Trees also were constructed using the Fitch–Margoliash algorithm (FITCH), unweighted parsimony (PAUP 3.1.1), and maximum likelihood quartet puzzling (PUZZLE 2.0). Multiple sequence alignments are available upon request.

Immunological Analysis of S9 and Acyl Carrier Protein (ACP). T. gondii rps9 and acpP EST clones (GenBank accession nos. 198724 and N60936, respectively) were cloned in-frame into pGEX expression vectors (Pharmacia) and glutathione S-transferase (GST)-fusion proteins expressed in Escherichia coli. Affinity-purified fusion proteins were used to immunize rabbits, and antibodies were affinity purified by using immobilized fusion protein. Western blot analysis was performed on detergent extracts of RH strain T. gondii tachyzoites, using secondary antibodies conjugated to alkaline phosphatase. For immunofluorescence, T. gondii tachyzoites were rapidly frozen, thawed in 3% paraformaldehyde in PBS, and post-fixed 20 min in cold methanol; antibody incubations were rapidly frozen, thawed in 3% paraformaldehyde in PBS, and post-fixed 20 min in cold methanol; antibody incubations were performed by using rabbit antiserum to ACP and affinity purified by using COLONTECH. Cells were counterstained with 5-6-carboxyfluorescein isothiocyanate-conjugated secondary antibodies. DNA counterstaining was carried out by using 5 μg/ml propidium iodide. Immunofluorescence of ultra-thin sections for electron microscopy was performed similarly, using parasites embedded in LR-Gold (London Resin) and secondary antibodies conjugated to 20 nm gold particles (BioCell). Cells were examined by using a Leica TCS 4D confocal microscope or a Philips CM120 BioTWIN transmission electron microscope. Controls using preimmune sera, secondary antibodies only, or anti-glutathione S-transferase antibodies were negative.

Green fluorescent Protein (GFP) Reporter Targeting. GFP fusion protein constructs were used by using T. gondii acpP-coding sequences corresponding to either the entire acpP gene product or the first 104 aa only. These fragments were amplified from tachyzoite cDNA by using sense primer 5’-gagaugggataagatGTCGACGAGAGATGCATCCCCGCAACGC-3’ and antisense primers 5’-TGAGACCTTGGCGTCGTAGGCCTCTGTTTTCTC-3’ or 5’-TGAGACCTTGGCGTCGTAGGCCTCTGTTTTCTC-3’ (for TgACPfull-GFP or TgACPleader-GFP, respectively) and ligated as BglII-AvrII fragments in place of the P30 sequences in plasmid pUlt30-GFP/sag-CAT (21). T. gondii tachyzoites were transfected and selected for stable transformants as described (21). Fluorescence was observed by using a Zeiss Axiovert 35-inverted microscope equipped with a 100 W Hg-vapor lamp. Western blotting and immunofluorescence of ACP–GFP fusion proteins was carried out as described (21) by using rabbit anti-GFP (CLONTECH) and anti-ACP, and colocalization with apicoplast DNA was examined by counterstaining with 4’,6-diamidino-2-phenylindole (3, 8).

**Thiolactomycin Growth Inhibition of T. falciparum.** Thiolactomycin was synthesized as described (22). Asynchronous cultures of P. falciparum-infected erythrocytes (strains D10 and W2 mef; 0.5% parasitemia) were inoculated into microtiter wells containing serial twofold dilutions of thiolactomycin in culture medium (final concentrations: 250–0.5 μg/ml = 1189–2.3 μM). 3H-labeled hypoxanthine was added in fresh hypoxanthine-free medium at 28 hr, and incorporation was assessed 18 hr later as described (23). Inhibition was plotted as the percentage of growth in drug-free medium (average of six replicates), and IC₅₀ values were estimated from this graph.

**RESULTS AND DISCUSSION**

Identification of Nuclear-Encoded Apicoplast Proteins. Searching of the T. gondii and P. falciparum databases for entries exhibiting similarity to nuclear-encoded genes that are known to be targeted to plastids in plant and algal systems (16) identified several candidate sequences, including ribosomal protein genes rps9 and rpl28 (encoding S9 and L28, respectively) and fatty acid biosynthetic genes acpP, fabH, and fabZ (encoding ACP, β-ke-taoacyl-ACP synthase III and β-hydroxyacyl-ACP dehydratase, respectively). The rps9 and rpl28, acpP, fabH, and fabZ genes are
undoubtedly nuclear because they (i) are absent from the apicoplast genomes (1, 3), (ii) were recovered from polyA+ cDNA, and (iii) harbor spliceosomal introns. Moreover, Southern analysis of *P. falciparum* pulsed field gel electrophoresis blots shows that acpP and fabH are located on chromosome 2 (data not shown). Molecular phylogenies for *rps9*, *rpl28*, and *fabH* revealed that these nuclear genes are unequivocally prokaryotic in nature and cluster with plastid and cyanobacterial homologs (Fig. 1). Phylogenetic analysis of ACP proteins also indicates a plastid origin for the *Toxoplasma* and *Plasmodium* acpP genes (data not shown), but these trees were less well resolved, due in part to the short protein length (~85 aa). β-hydroxyacyl-ACP dehydratase (FabZ) is present in plant chloroplasts (24) but has not been cloned yet from plants or algae, preventing informative phylogenetic analysis.

These data suggest that all of the above genes now present in the nucleus of apicomplexan parasites derive from the endosymbiont. (either directly from the plastid genome or from the endosymbiont nucleus). It is noteworthy that (β-ketoacyl-ACP synthase III) FabH, ACP, L28, and S9 are plastid encoded in red algae but nuclear encoded in plants (16). A corollary of such intracellular gene relocation is that protein products of the nuclear genes acquired from the endosymbiont are likely to be targeted back to the organelle of origin. This is especially likely for *rps9* and *rpl28*, which are not encoded in the apicoplast genome of *Toxoplasma* (3) or *Plasmodium* (1), but which are likely to be required for protein translation.

Antisera raised against recombinant *Toxoplasma rps9* and *acpP* gene products were used to test for apicoplast targeting of these proteins in *Toxoplasma* tachyzoites. As shown in Fig. 2, immunolocalization of S9 and ACP (respectively) in intact *T. gondii* cells demonstrates that these proteins are restricted to a distinct region of the parasite similar to the location of the apicoplast. Some cells show two plastids, which are probably division stages. (C) Counterstaining with propidium iodide (red) confirms colocalization of ACP (green) with the extranuclear apicoplast DNA (Inset shows DNA staining only; arrowhead indicates apicoplast DNA). (D and E) Detection of ACP by immunogold labeling of ultrathin sections shows strong labeling of the apicoplast (arrowhead). Nu, nucleus; Mi, mitochondrion; Go, Golgi apparatus. (F) The N-terminal domain of ACP (TgACPLeader-GFP) is sufficient to target GFP to the apicoplast, and the recombinant protein can be visualized in living cells. (G–I) In fixed cells labeled with anti-GFP (green) and counterstained with DAPI (blue), TgACPLeader-GFP can be seen to colocalize with the apicoplast DNA (arrowheads indicate apicoplasts in two cells). Color images were collected independently and overlaid on top of phase-contrast micrographs. (White scale bars = 10 µm, black scale bars = 200 nm).
Primary Endosymbiosis describes the uptake of a prokaryote by a eukaryote. Plastids derived by primary endosymbiosis are generally surrounded by two membranes, and targeting of nuclear-encoded gene products to the endosymbiont is effected by an N-terminal transit peptide (T). Secondary endosymbiotic plastid origin involves a heterotrophic eu-karyote phagocytosing a photosynthetic eukaryote produced by primary endosymbiosis. The secondary endosymbiont’s cytoplasm and nucleus (N”) are typically lost or heavily reduced, and the resulting plastid is surrounded by four membranes, the outermost of which derives from the phagocytic membrane. Sometimes one of the two outer membranes is lost at this point, resulting in a total of three. Targeting of nucleus-encoded (N”) gene products to secondary plastids requires a signal peptide (S) to mediate protein passage across the outer membrane(s) followed by a transit peptide (T) for import across the inner membrane pair.

Nuclear-Encoded Apicoplast Proteins Contain Cleavable Pre-sequences that Mediate Targeting. Plastid targeting can be classified into two broad categories depending on evolutionary origin and ultrastructure, particularly the number of membranes surrounding the plastid. In plants, green algae, red algae, and glaucophytes, whose plastids are enclosed by two membranes, targeting is mediated by an N-terminal presequence known as the transit peptide, which is removed after import (Fig. 3) (25). Plant and green algal transit peptides are the best characterized of these, and although no consensus sequence or secondary structure is evident, they are typically 25–125 aa in length, basic, and rich in serine and threonine (25, 26).

Targeting presents a more complex problem in organisms with more than two membranes around the plastid. Euglenoids and dinoflagellates have three plastid membranes, and certain other algal groups (heterokont and haptophyte algae, chlorarachniophytes, and cryptomonads) have four plastid membranes. Plastids with four membranes were almost certainly acquired by secondary endosymbiosis, in which a phagotrophic eukaryote engulfed (or was invaded by) a plastid-containing eukaryote (Fig. 3) (12). Plastids with three membranes may have been acquired either by secondary (12) or primary (14) endosymbiosis. The details of protein trafficking to plastids containing multiple membranes is not well understood, but because the outermost membrane is part of the host’s endomembrane system, targeting apparently commences via the secretory pathway into the endoplasmic reticulum (ER) courtesy of a classic signal peptide (27–29). Subsequent targeting across the inner pair of plastid membranes involves a downstream transit peptide (Fig. 3) (27–29).

The nuclear-encoded apicoplast genes identified here are all predicted to encode substantial N-terminal extensions when compared with the equivalent plastid and bacterial proteins (Fig. 4). The extreme N-terminal regions (16–34 aa) of these extensions resemble classic signal peptides, containing a hydrophobic domain followed by a “von Heijne” cleavage site (30). Downstream of the predicted signal peptides, the Toxoplasma proteins exhibit the general features of transit peptides: net positive charge and enrichment for serine and threonine residues. Putative Toxoplasma transit peptides range from 57–107 aa in length. The putative Plasmodium transit peptides also have a net positive charge, are somewhat shorter (30–42 aa), and rich in lysine and asparagine, probably reflecting the strong A/T bias of the Plasmodium genome (>80%) as well as the apparently loose constraints of transit peptides (26). No apicoplast transit peptide cleavage motif has been identified to date. Several targeting peptide boundaries are located near introns (Fig. 4), possibly indicating acquisition of these motifs by exon shuffling (31). It is also interesting to note that the locations of two introns are shared in T. gondii and P. falciparum acpP genes (Fig. 4), suggesting these genes are orthologous and that the gene transfer from endosymbiont to parasite nucleus and acquisition of targeting sequences occurred prior to the divergence of these species.

Processing of nuclear-encoded apicoplast proteins was examined by Western blot analysis, as shown in Fig. 5. S9 and ACP antisera recognized protein bands of the expected size for the mature proteins (lower bands in Fig. 5B), suggesting that the predicted N-terminal extensions are removed. Interestingly, both antisera also recognize larger proteins corresponding to the predicted size of the relevant mature protein plus transit peptide in each case. These larger bands probably represent processing intermediates associated with apicoplast targeting, either in the process of traversing the multiple apicoplast membranes, or awaiting proteolytic cleavage in the apicoplast stroma. As signal peptides are usually removed during cotranslational import into the ER, it is not surprising that full length precursors including the signal peptide were not detected.

To test the hypothesis that the N-terminal extensions present on nuclear-encoded apicoplast proteins mediate targeting, GFP fusion proteins were constructed as shown in Fig. 5A by using either the entire T. gondii acpP gene or only the putative N-terminal targeting information (signal and transit peptides). These constructs, designated TgACPleader-GFP and TgAPCleader-GFP, respectively, were transfected into T. gondii. As indicated in Fig. 5F–I, apicoplast-specific targeting was observed in both transient and stable transfectants, using both constructs. Successful targeting of the truncated ACP fusion protein (TgACPleader-GFP) demonstrates that the 104-aa N-terminal presequence of TgACP is sufficient to target a protein into the apicoplast. Interestingly, Western blot analyses of GFP fusion proteins by using anti-GFP antisera reveal both processed and unprocessed pools of apicoplast-targeted proteins of the ex-
system, coupled with the accessibility of T. gondii to molecular genetic manipulation (21), provides a convenient means to dissect plastid targeting in a secondary endosymbiont.

By using this targeting model, we have identified three additional Plasmodium genes whose products are likely to be targeted to the apicoplast: cdpP (a plastid protease GenBank accession no. AL009009), EF-TS (elongation factor TS, Sanger ID MP03010), and cpm60 (a plastid chaperonin GenBank accession no. X75420). Conversely, we can now predict proteins not likely to be directed to the apicoplast. For example, despite previous speculation that the apicoplast might harbor a shikimate pathway for essential amino acid synthesis (7, 10, 32), the lack of N-terminal extensions on both P. falciparum and T. gondii genes encoding chorismate synthase (32) suggests a cytoplasmic localization, as found in fungi (33). Characterization of precursors thus provides a powerful predictive tool for compiling a picture of apicoplast proteins and metabolic pathways from the ever-expanding genomes databases.

**Does the Apicoplast Synthesize Fatty Acids?** Identification of nuclear-encoded apicoplast proteins with unequivocal similarity to the fatty acid biosynthetic genes acp, fabH, and fabZ offers a new insight into the metabolic activity of this enigmatic organelle. By analogy with plant and algal plastids, apicoplasts have been suggested to perform three possible anabolic pathways: essential amino acid biosynthesis, heme biosynthesis, and fatty acid biosynthesis (1, 3, 7, 10). Although essential amino acid biosynthesis and heme biosynthesis in apicoplasts cannot yet be discounted, available evidence suggests these functions are fulfilled by cytosolic pathways (refs. 1, 3, 7, 15 and see above). In contrast, the presence of ACp and (very probable) FabH and FabZ in the apicoplast indicates that at least some fatty acid biosynthesis occurs in the apicoplast. All three proteins are members of the fatty acid synthase multi-enzyme complex (24, 34). ACp plays a central role in fatty acid biosynthesis by holding the forming acyl chain, whereas FabH and FabZ are involved in the condensation and dehydration steps, respectively, of acetyl addition during acyl chain elongation (24, 34). Partial sequence also has been obtained for another fatty acid synthase component from P. falciparum, fabF, which encodes β-ketoacyl-ACP synthase II (34), further implicating a fatty acid biosynthesis pathway in these parasites.

There are two types of fatty acid biosynthesis. Type I is found in the cytosol of animals and fungi (35). Type II is widespread among bacteria but in eukaryotes is restricted to the plastids of plants and algae (24, 34) and perhaps the mitochondria of yeast (36), which reflects the origin of these organelles from endosymbiotic bacteria (16). The T. gondii and P. falciparum fatty acid biosynthesis genes characterized here are of type II, which is consistent with localization of the protein products to the apicomplexan plastid. Importantly, the different types of fatty acid biosynthesis allow for selective perturbation. The antibiotic thiolaactomycin is a selective inhibitor of type II fatty acid biosynthesis (37). In E. coli, thiolaactomycin inhibits the condensing enzymes FabB, FabF, and FabH (34, 38), and it inhibits the equivalent plastid enzymes in plants (24, 39). In contrast, thiolaactomycin has no effect on fatty acid biosynthesis (type I) of Saccharomyces cerevisiae, Candida albicans, and rat liver (37). We examined the effect of thiolaactomycin on the malaria parasite. In vitro growth inhibition assays using a racemic mixture of (R and S) thiolaactomycin with Plasmodium falciparum show an IC50 of ~50 μM (Fig. 6). This level of inhibition is comparable with that seen for fatty acid biosynthesis in isolated pea and spinach plastids (IC50 of 10–100 μM; refs. 40 and 41), as well as for the individual condensing enzymes (5–>25 μM) (24, 39). Thiolaactomycin inhibition of malaria thus provides additional supporting evidence for a type II fatty acid biosynthetic pathway in apicoplasts. Proof of this hypothesis awaits biochemical demonstration of fatty acid production in the apicoplast, perhaps by demonstrating presence of type II signature lipids (24, 42) in apicomplexa.

Why would the apicoplast synthesize lipids? Plant plastids require unusual fatty acids (43), and it is possible that apico-
plast fatty acid biosynthesis merely provides an extended house-keeping function for the organelle. However, in plants and certain algae, the plastid exports lipids to the rest of the cell and is the sole site of fatty acid biosynthesis (24, 42). A fatty acid, or lipid, could thus be the mysterious “factor X” proposed to be produced by Toxoplasma apicoplasts (44), and perhaps such lipids are an essential component for parasitophorous vacuole formation because poisoning the apicoplast inhibits parasite replication at an early stage after infection (8).

Concluding Remarks. We have characterized the first nuclear-encoded plastid proteins in apicomplexan parasites. We have demonstrated that an N-terminal extension on these proteins mediates targeting, probably via the endomembrane system. Several of the proteins are members of the type II fatty acid biosynthesis complex; a pathway not recognized previously in these parasites. Preliminary trials with an agent known to perturb the type II pathway suggests that fatty acid biosynthesis may be an excellent new target for combating apicomplexan-mediated diseases.

We are grateful to J. W. Ajikoe for a T. gondii RH cDNA library, the Wash-U-Merk Toxoplasma Expressed Sequence Tag Project for T. gondii Expressed Sequence Tag clones, M. E. Wickham for P. falciparum cDNA, M. Guy for thiolactomycin synthesis, V. Su and M. Reed for technical assistance, J. Thompson for P. falciparum Southern analysis, M. Reith forinformatics assistance, and A. Stoltzfus for MOL2CON. This research was supported by awards from the National Institute of Allergy and Infectious Diseases and the U.S. Department of Defense.