Processing of an Apicoplast Leader Sequence in *Plasmodium* falciparum and the Identification of a Putative Leader Cleavage Enzyme*

Received for publication, February 20, 2002, and in revised form, April 23, 2002 Published, JBC Papers in Press, April 24, 2002, DOI 10.1074/jbc.M201748200

Giel G. van Dooren‡, Vanessa Su, Marthe C. D'Ombrain, and Geoffrey I. McFadden§

From the Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Melboure 3010, Australia

The plastid (apicoplast) of the malaria-causing parasite Plasmodium falciparum was derived via a secondary endosymbiotic process. As in other secondary endosymbionts, numerous genes for apicoplast proteins are located in the nucleus, and the encoded proteins are targeted to the organelle courtesy of a bipartite N-terminal extension. The first part of this leader sequence is a signal peptide that targets proteins to the secretory pathway. The second, so-called transit peptide region is required to direct proteins from the secretory pathway across the multiple membranes surrounding the apicoplast. In this paper we perform a pulse-chase experiment and N-terminal sequencing to show that the transit peptide of an apicoplast-targeted protein is cleaved, presumably upon import of the protein into the apicoplast. We identify a gene whose product likely performs this cleavage reaction, namely a stromal-processing peptidase (SPP) homologue. In plants SPP cleaves the transit peptides of plastid-targeted proteins. The P. falciparum SPP homologue contains a bipartite N-terminal apicoplast-targeting leader. Interestingly, it shares this leader sequence with a Δ -aminolevulinic acid dehydratase homologue via an alternative splicing event.

Plasmodium spp., the causative agents of malaria, belong to a family of intracellular parasites called the Apicomplexa. *Plasmodium* infects approximately 300 million people annually, causing over 1 million deaths, the great majority of which are caused by one species, *Plasmodium falciparum* (1). *P. falciparum* infects both humans and mosquitoes during its life cycle, with the pathogenic part of this cycle occurring predominantly in the erythrocytes of humans. The discovery of a non-photosynthetic plastid (the apicoplast) in the Apicomplexa has opened up a new area of anti-malarial drug targets (2, 3). However, the rational development of drugs targeting the apicoplast requires knowledge of apicoplast function. Preliminary studies indicate that the apicoplast is a site of fatty acid and isoprenoid biosynthesis (4–6), and drugs targeting these pathways have been shown to kill *P. falciparum* (4, 6, 7).

Like plant plastids, the apicoplast contains a reduced bacterial-like genome (8), from which a small number of proteins are

expressed. The great majority of apicoplast proteins, as in plant plastids, are encoded in the nucleus and must be post-translationally targeted to the plastid. In plants, nuclear-encoded plastid proteins require a cleavable, N-terminal sequence called the transit peptide, which directs these proteins across the two membranes surrounding plant plastids (for reviews, see Refs. 9 and 10). Once in the plastid stroma, this transit peptide is cleaved by a stromal-processing peptidase (SPP¹; Refs. 11 and 12). Apicoplasts, however, are bound by four membranes (Ref. 2, but see Ref. 13), and proteins targeted to this organelle have been shown to require a bipartite N-terminal leader sequence (5, 14, 15). By fusing these N-terminal leader sequences to green fluorescent reporter protein (GFP), Waller *et al.* (5) show that the first part of this leader functions as a signal peptide, directing proteins to the secretory pathway. The second part of the leader is necessary to direct these proteins away from the default secretory pathway and into the apicoplast (5). Indeed, this second region has several similarities to the transit peptides of plants, and the transit peptidelike region of an apicoplast-targeted protein from the related species Toxoplasma gondii is sufficient to target proteins into isolated pea chloroplasts (14).

In a Western blot analysis of these apicoplast-targeted proteins, Waller et al. (5) found a prominent band whose size corresponded to that of the mature protein as well as a less abundant band slightly greater in size. In this paper, we perform a pulse-chase experiment to demonstrate that this larger band represents the apicoplast-targeted protein before cleavage of the transit peptide-like region has occurred. We then identify the site of transit peptide cleavage by N-terminal sequencing of the processed protein. Finally, we identify a putative SPP homologue in P. falciparum, which we propose is responsible for this cleavage. Interestingly, this gene shares, through alternative splicing, a putative N-terminal apicoplasttargeting sequence with a Δ -aminolevulinic acid dehydratase (ALAD; also known as porphobilinogen synthase; EC 4.2.1.24) homologue, a protein that functions in heme biosynthesis. The evolution of this alternative splicing is considered.

EXPERIMENTAL PROCEDURES

Transfected Cell Line—Pulse-chase and N-terminal sequencing was performed on *P. falciparum* cells transfected with the acyl carrier protein (ACP) leader-GFP fusion protein, produced by Waller *et al.* (5). The fusion protein consists of the bipartite N-terminal leader sequence of the ACP (a fatty acid biosynthesis protein) fused to the N terminus of GFP. Cells were cultured using standard techniques (16), with py-

^{*} The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[‡] Supported by an Australian Postgraduate Award.

[§] Supported by the Australian Research Council, the National Health and Medical Research Council, and a Howard Hughes Medical Institute International Research Scholar. To whom correspondence should be addressed. Tel.: 61-3-8344-5054; Fax: 61-3-9347-1071; E-mail: gim@ unimelb.edu.au.

¹ The abbreviations used are: SPP, stromal-processing peptidase; GFP, green fluorescent protein; ALAD, Δ-aminolevulinic acid dehydratase; CAPS, 3-(cyclohexylamino)propanesulfonic acid; ACP, acyl carrier protein; MPP, mitochondrial-processing peptidase; PfSPP, *P. falciparum* SPP; PfMPPA, *P. falciparum* α-subunit of MPP; PfMPPB, *P. falciparum* β-subunit of MPP; PfALAD, *P. falciparum* ALAD.

rimethamine drug added to select for parasites containing the ACP leader-GFP-containing episomes (5).

Pulse-chase of the ACP Leader-GFP Construct-Red cell cultures containing parasites were washed twice in methionine-free RPMI-HEPES medium (Sigma) then incubated for 45 min in RPMI-HEPES medium containing 35 S-labeled methionine and cysteine (PerkinElmer Life Sciences Expre 35 S 35 S protein labeling mix, final concentration 80 μ Ci/ml) supplemented with 5% Albumax (Invitrogen) and 5% heatinactivated human serum at 37 °C and under 5% CO₂ and 1% O₂ gas in nitrogen. Cells were pelleted and washed twice with non-radioactive RPMI-HEPES (Sigma) and resuspended in this medium supplemented with 5% Albumax (Invitrogen) and 5% heat inactivated human serum, and equal aliquots were pipetted into 5 separate dishes. Cells were kept at 37 °C under 5% $\rm CO_2$ and 1% $\rm O_2$ gas in nitrogen and harvested at t=0, 1, 2, 3, and 4 h by saponin lysis and subsequent immunoprecipitation (see below). Saponin lysis involves selectively permeabilizing the red cell membrane and parasitophorous vacuole with 0.15% (w/v) saponin in phosphate-buffered saline (0.8% w/v NaCl, 0.02% w/v KCl, 0.14% w/v Na₂HPO₄, 0.02% KH₂PO₄, pH 7.4). After immunoprecipitation, samples were boiled in sample buffer, and separated by SDS-PAGE (17). Radioactive bands were visualized using x-ray imaging film (Eastman Kodak Co., Biomax).

Immunoprecipitation of ACP Leader-GFP Constructs—After saponin lysis (see above), parasite-containing pellets were washed in phosphatebuffered saline and lysed in 500 μ l of immunoprecipitation lysis buffer (0.05 M Tris-HCl, pH 7.5, 1% v/v Triton X-100, 0.6 M KCl) containing protease inhibitors (0.4 mM phenylmethylsulfonyl fluoride, 5 μ g/ml pepstatin A, 5 μ g/ml amino-n-caproic acid, 5 μ g/ml aprotinin, 5 μ g/ml leupeptin). After lysis at room temperature for 5 min and then on ice for 30 min, the cell extract was spun down, and the supernatant was collected.

To remove proteins that bind non-specifically to protein A or the Sepharose beads used in the immunoprecipitation, the supernatant was incubated for 1 h in 80 μ l of protein A-Sepharose CL-4B beads (Amersham Biosciences) made to a 50% (v/v) slurry with immunoprecipitation wash buffer (0.05 M Tris-HCl, pH 7.5, 1% v/v Triton X-100, 1 mM EDTA, 0.15 M NaCl, 0.25% w/v bovine serum albumin). 5 μ l of polyclonal rabbit anti-GFP antibody (CLONTECH, catalog number 8372) was added to 60 μ l of 50% Sepharose slurry. This was incubated for 1 h at 4 °C, added to the pre-cleared supernatant from the previous step, and incubated for a further 2 h at 4 °C. Beads were washed four times in wash buffer and twice in phosphate-buffered saline. Proteins were eluted from the beads by boiling in non-reducing sample buffer (18).

N-terminal Sequencing—Immunoprecipitated proteins were separated using SDS-PAGE (17) and transferred onto polyvinylidene difluoride-plus membranes (Micron Separations Inc.; manufacturer's protocol) in transfer buffer (10% v/v methanol, 10 mM CAPS, pH 11). A band corresponding to the expected size for processed GFP was visualized with a Coomassie Blue stain, excised, and subjected to N-terminal sequencing on a Beckman N-terminal sequencer.

Gene Cloning and Sequencing-Homologues of a stromal-processing peptidase (PfSPP), an α -subunit mitochondrial-processing peptidase (PfMPPA), a β -subunit mitochondrial-processing peptidase (PfMPPB), and a Δ -aminolevulinic acid dehydratase (PfALAD) were identified in the P. falciparum genome data base.² The preliminary sequence data for P. falciparum chromosome 14 was obtained from The Institute for Genomic Research web site (www.tigr.org). Sequencing of chromosome 14 was part of the International Malaria Genome Sequencing Project and was supported by awards from the Burroughs Wellcome Fund and the United States Department of Defense. Sequence data for P. falciparum chromosomes 6, 7, 8, and 9 was obtained from The Sanger Institute web site at www.sanger.ac.uk/Projects/P_falciparum. Sequencing of P. falciparum chromosomes 6, 7, 8, and 9 was accomplished as part of the Malaria Genome Project with support by The Wellcome Trust. Reverse transcription-PCR using gene-specific primers was used to construct cDNA clones containing the full protein coding region for each of these genes. GenBankTM accession numbers of the genes are as follows: PfSPP (AF453250), PfMPPA (AY069958), PfMPPB (AY064478), and PfALAD (AY064477). cDNA was produced from the D10 line of P. falciparum parasites. To determine whether any of these sequences has an N-terminal, apicoplast-targeting leader, we used the Prediction of Apicoplast Targeted Sequences (PATS) neural network (modlab.biologie.uni-freiburg.de/gecco/pats) (19).

Sequence Analysis-Although there is probably more than one sort of SPP in plants (20), only one has been cloned (21). This SPP belongs to a family of metalloendopeptidases known as the pitrilysin family (Ref. 21; peptidase family M16). To determine whether any of the pitrilysin proteins in P. falciparum are homologous to the plant SPPs, a phylogenetic tree of the pitrilysin family was constructed. The P. falciparum pitrilysin homologues (PfSPP, PfMPPA, PfMPPB, and falcilysin, an enzyme involved in the hemoglobin degradation pathway (accession number AAF06062)) were translated to the amino acid sequence and aligned to other pitrilysin protein sequences retrieved from the NCBI data base using ClustalW 1.8 and PIMA 1.4 in BCM Search Launcher $(dot.imgen.bcm.tmc.edu: 9331/multi-align/multi-align.html). \ Align-indication (dot.imgen.bcm.tmc.edu: 9331/multi-align/multi-align/multi-align.html).$ ments were refined manually in PAUP*4.0b3. Graphical representations of alignments were produced in Boxshade 3.21 (www.ch.embnet.org/software/BOX_form.html). The other pitrilysins used in this analysis were Escherichia coli pitrilysin (P05458), an E. coli putative zinc protease (P31828), a Borrelia burgdorferi putative zinc protease (G70166), two putative zinc proteases from Synechocystis (S74378 and S77157), human insulysin (NP_004960), rat nardilysin (P47245), Arabidopsis thaliana SPP (BAB10480), Pisum sativum SPP (AAA81472), *Eimeria bovis* sporozoite developmental protein (SDP: P42789), A. thaliana \beta-MPP (AAF14827), rat β-MPP (Q03346), Schizosaccharomyces pombe β -MPP (T42428), A. thaliana α -MPP (AAK59675), rat α -MPP (P20069), and S. pombe α -MPP (CAA22672). In addition to those available in GenBankTM, two sequences representing homologues of PfSPP in Theileria parva and Plasmodium yoelii were included in the analysis. Preliminary sequence data from the P. yoelii genome was obtained from The Institute for Genomic Research web site (www.tigr.org). This sequencing program is carried on in collaboration with the Naval Medical Research Center and is supported by the United States Department of Defense. Preliminary sequence data from the T. parva genome was obtained from The Institute for Genomic Research web site at www. tigr.org. Neighbor-joining and maximum parsimony trees were constructed with PAUP*4.0b3. Trees excluded regions of the alignment where poor matching occurred, taking into account the common region of pitrilysin proteins, which centers around the active site area, leaving 189 characters for analysis. Bootstrap values from 1000 replicates were obtained.

Northern Analysis of PfSPP and PfALAD-Sequencing of the PfSPP and PfALAD cDNAs and comparison with the genome data revealed that the two 5' exons incorporated into the two cDNAs are identical, suggesting that the N-terminal region of both proteins are identical. To begin analyzing the nature of this shared N terminus, a Northern analysis was performed. RNA was extracted as described (22) and electrophoretically separated on a 0.9% agarose gel containing 5 mM guanidine thiocyanate followed by capillary transfer to Hybond N+ membrane (Amersham Biosciences) overnight (22). Membranes were prehybridized for 3 h in 7% SDS, 0.5 M sodium phosphate, 1 mM EDTA. Three probes were used for the analysis, one binding to the shared 5' region (amplified by the primers 5'-CATTATGAAAATTTTGAAAATTT-TCAG and 5'-CACGGGATTATTCCATCTTC), one specific to PfALAD (5'-GGGAAAGACGTATAAAGAGG and 5'-CGTTATAAACAGCAAT-AGG), and one specific to PfSPP (5'-GAGCTACAAAATGGGTTAAG-CAAC and 5'-CGTCTGTGTCTTTAAACG). Probes were labeled with 25 µCi of [³²P]dCTP following the Prime-a-Gene (Promega) manufacturer's protocol, and membranes were hybridized overnight in the prehybridization buffer at 65 °C. Membranes were washed 2–4 times in 0.5–1 imesSSC ($1 \times SSC = 0.15$ M NaCl and 0.015 M sodium citrate), 0.1% SDS at 60-65 °C depending on the probe used, and the film was developed at -70 °C for up to 2 weeks. Bands were sized using a 0.28-6.58-kb RNA marker (Promega; 3 μ l in 30% v/v formamide, 4% formaldehyde, 1× RNA gel-loading buffer (18) in 1 × Tris-borate/EDTA (Merck)).

RESULTS

Pulse-chase to Analyze Processing of Apicoplast-targeted GFP—Proteins targeted to plant and green algal plastids contain N-terminal transit peptides that are both necessary and sufficient to target those proteins into the plastid (23). After import into the plant plastid, this transit peptide is cleaved (11). Proteins targeted to the apicoplast of Apicomplexa also contain N-terminal extensions, which differ from plant counterparts in that they consist of two parts. The first part is a signal peptide that directs proteins to the secretory pathway and is probably cleaved upon the protein entering the endo-

² Sequencing of the *Plasmodium falciparum*, *P. yoelli*, and *Theleria parra* genomes was supported by awards from the Burroughs Wellcome Fund, the United States Department of Defense, The Wellcome Trust, International Livestock Research Institute, and The Institute for Genomic Research.



FIG. 1. A pulse-chase experiment demonstrating post-translational processing of an apicoplast-targeted GFP fusion protein. P. falciparum cells were pulsed with radioactive methionine and cysteine for 45 min. Cells were harvested at 1-h intervals over 4 h. GFP was purified by immunoprecipitation and separated by SDS-PAGE. The band at around 28 kDa represents processed GFP, with the band slightly higher than this thought to represent GFP with the transit peptide still attached (TP-GFP). Directly after the radioactive pulse (t =0), the relative amount of TP-GFP is high compared with processed GFP. Over the course of the chase, TP-GFP decreases in relative intensity, corresponding to cleavage of the transit peptide region to form processed GFP. After the 45-min pulse, there is a small amount of processed GFP. Assuming the processing occurs in the apicoplast, this suggests that 45 min closely approximates the minimum amount of time it takes for import of cytosolically translated, apicoplast-targeted proteins into the apicoplast.

plasmic reticulum (5). The second part of the leader (the socalled transit peptide) is thought to direct proteins through the secretory pathway and into the multi-membraned apicoplast. Western blots of apicoplast-targeted proteins have consistently revealed the presence of two forms of the given apicoplast protein (4, 5, 24). It is thought that the smaller of these two forms represents the mature, active protein, whereas the other, larger protein represents the pre-processed form, which still bears the transit peptide part of the N-terminal leader. To determine whether this processing occurs and to analyze the kinetics of this process, a pulse-chase experiment was performed on a *P. falciparum* cell line containing an apicoplasttargeted GFP construct.

An anti-GFP antibody was used to immunoprecipitate the variously processed forms of GFP in the ACP leader-GFP cell line produced by Waller et al. (5). Western blotting was used to confirm that the procedure was successful in immunoprecipitating GFP (results not shown). Cells were pulsed for 45 min with radioactive methionine and cysteine. Immunoprecipitated proteins obtained from cells harvested at hourly intervals after the pulse were separated using SDS-PAGE, and an autoradiograph image was taken (Fig. 1). The two protein bands at around 28 and 32 kDa correspond to the expected sizes of the processed and transit peptide-containing ACP leader-GFP proteins, respectively (Fig. 1). Over the course of the chase, the top band became reduced in relative intensity compared with the bottom band. After the 45-min pulse (at t = 0), some of the bottom band was already present, increasing in relative intensity over the course of the chase. This suggests that the minimum time for targeting proteins to the apicoplast is less than 45 min. It is interesting that the total amount of immunoprecipitated protein at t = 0 in our pulse-chase experiment seems to be less than the other time points. We interpret this as resulting from some radioactive label being present in the red cell and parasite that, after the 45-min pulse, was not yet incorporated into protein.

N-terminal Sequencing—To confirm the identity of the immunoprecipitated protein and to determine whether and where cleavage of the transit peptide-like domain occurs, we at-



FIG. 2. The transit peptide cleavage site of the ACP leader-GFP fusion protein. The transit peptide cleavage site of the ACP leader-GFP fusion protein was determined by N-terminal sequencing of protein purified by immunoprecipitation. This diagram indicates where this cleavage site occurs. The five amino acids (LNRKN) determined by N-terminal sequencing of the 28-kDa band in Fig. 1 are shown in *bold*. Also shown is the putative signal peptide cleavage site, as predicted by SignalP V2.0.b2 (www.cbs.dtu.dk/services/SignalP-2.0). Cleavage of the signal peptide probably occurs shortly after co-translational import into the endoplasmic reticulum. It is thought that the transit peptide (*lowercase*) then mediates transportation of the apicoplast-targeted protein through the secretory pathway and across the membranes surrounding the apicoplast. Transit peptide cleavage most likely occurs after the protein has been imported into the apicoplast.

tempted to sequence the N terminus of the processed form of the apicoplast-targeted GFP construct. Immunoprecipitated proteins from the ACP leader-GFP cell line were separated using SDS-PAGE and transferred to the polyvinylidene difluoride membrane. A band of around 28 kDa, corresponding to the expected size for processed GFP, was visualized after Coomassie staining (not shown). N-terminal sequencing of this band from the polyvinylidene difluoride membrane produced a five-residue sequence of LNRKN, corresponding to the first five amino acids of the mature ACP protein. This sequence begins 25 amino acids downstream of the predicted cleavage site for the signal peptide component of the bipartite leader of ACP (Fig. 2). The pre-processed form of GFP was not present in great enough abundance to enable N-terminal sequencing.

Phylogenetic Analysis of PfSPP-In plants, a SPP cleaves plant transit peptides soon after plastid-targeted proteins are imported into the plastid (12, 25, 26). The stromal-processing peptidases of plants belong to a family of metalloendopeptidases known as the pitrilysins (21), which also include both subunits of mitochondrial-processing peptidases (27) and the insulin-degrading enzyme insulysin (28). To assess the phylogeny of the cloned PfSPP homologue within the pitrilysin family, a phylogenetic tree incorporating a range of pitrilysin proteins was constructed (Fig. 3). Bootstrap values for neighbor joining (93%) and maximum parsimony (85%) trees support the grouping of PfSPP as well as the other apicomplexan SPP homologues with SPPs of plants. Contrary to previous suggestions (29), there is no obvious Synechocystis SPP homologue, with both Synechocystis pitrilysin proteins included in this tree not grouping with the SPPs. Of the other P. falciparum pitrilysins sequenced in this paper, one (PfMPPB) formed a clade with the β -MPP subunits of other organisms, whereas the other (PfMPPA) grouped with the α -MPP subunits of other organisms.

All catalytically active members of the pitrilysin family contain several highly conserved residues (30), including a conserved histidine-X-X-glutamate-histidine (HXXEH) zinc binding domain (28, 31). Sequencing of cDNA from the putative SPP homologue reveals that it contains this domain as well as sharing several other highly conserved residues characteristic of the pitrilysin family (Fig. 4). Active pitrilysins contain a conserved glutamate downstream of the HXXEH motif that functions in the active site (32) and a second downstream glutamate that functions in zinc binding (33). Pitrilysins also contain a conserved asparagine around 25 residues downstream of the HXXEH motif and a tyrosine residue around 150



FIG. 3. Phylogenetic tree of the pitrilysin family. An unrooted neighbor-joining tree is shown, with bootstrap confidence values (1000 replications) shown for both neighbor-joining (left) and maximum parsimony (right) algorithms. Bootstrap values of less than 50% are indicated by a dash (-). The putative P. falciparum SPP and the corresponding homologues in P. yoelii and T. parva group with the SPPs of plants, with bootstrap values of 93% for neighbor-joining and 85% for maximum parsimony algorithms. The SPP group appears to branch with putative zinc proteases from E. coli and B. burgdorferi and not with any Synechocystis pitrilysin proteins as previously predicted (29). MPP group together with reasonable bootstrap support, with the α and β subunits forming apparent clades within this group. Falcilysin, a P. falciparum food vacuole protein, appears quite distinct from other members of the group represented on the tree. E. coli pitrilysin clades with several eukaryotic proteins, including human insulysin and E. bovis sporozoite developmental protein (SDP). The tree is based on 189 characters

residues downstream of the HXXEH motif (30). An alignment of PfSPP to other pitrilysins (Fig. 4) indicates that PfSPP contains all of these conserved residues and motifs except the conserved tyrosine residue, with PfSPP having a Phe in this position. One other catalytically active pitrilysin, the AXL1 protein involved in yeast budding (34), resembles PfSPP in having a phenylalanine residue in this position, suggesting that either tyrosine or phenylalanine (both aromatic residues) may be sufficient. Richter and Lamppa (12) identify four regions in plant SPPs that share a high level of identity. In an alignment of PfSPP to plant SPPs (not shown), PfSPP matches well to plant SPPs in region I (32% identity, 53% similarity), which encompasses the catalytic site and zinc binding domain. Nothing is known about the function of the other regions, but PfSPP matches less well to plant SPPs in regions II (24% identity, 44% similarity) and IV (19% identity, 38% similarity) and poorly to region III (11% identity, 26% similarity).

PfSPP Shares an N-terminal Leader with a Heme Biosynthesis Protein—The phylogenetic tree of pitrilysin proteins suggests PfSPP is a homologue of plant SPPs and, therefore, a likely candidate for cleavage of the transit peptide-like domain of apicoplast-targeted proteins. The transit peptides of apicoplast-targeting leader sequences are required to direct proteins from the secretory pathway to within the apicoplast (5). Thus, one prerequisite of PfSPP functioning in cleaving transit peptides of apicoplast-targeted proteins is that this nuclear-encoded protein is itself targeted to the apicoplast. All apicoplasttargeted proteins identified to date contain a bipartite leader

sequence (4-6, 14, 24, 35). In P. falciparum, this leader consists of a hydrophobic signal peptide at the very N terminus followed by a region rich in basic and asparagine residues (19).³ Sequencing of PfSPP cDNA and comparison with the genome data revealed that it encodes a large transcript consisting of six exons. The N terminus of the predicted protein sequence contains a hydrophobic region followed by a region rich in basic and asparagine residues, consistent with PfSPP being an apicoplast-targeted protein. The PATS neural network, designed specifically to detect apicoplast-targeted proteins in P. falciparum (19), predicts PfSPP to be apicoplast-targeted. Unexpectedly, however, the two 5' exons of PfSPP (*E1* and *E2* in Fig. 5), which encode most of the predicted apicoplast leader sequence, are not immediately upstream of the exons comprising the remainder of the SPP gene. Rather, these exons are immediately upstream of another gene, PfALAD, situated 5' to the remaining SPP exons (Fig. 5). This suggested that the two genes might share the leader sequence exons by alternative splicing (Fig. 5). PfALAD, a ALAD homologue (36), is involved in the heme biosynthesis pathway, and in plants ALAD is located in the plastid (37). Sequencing of a cDNA clone of PfALAD confirms that it also bears the same two 5' exons, encoding the same bipartite apicoplast-targeting domain found on PfSPP. The ALAD and SPP cDNA sequences are incomplete at their 5' and 3' ends, meaning that the exact sizes of E1, E6, and E10 (Fig. 5) are not known. A recent paper sequenced the entire ALAD transcript and found it to be around 1.6 kb in size (38). Contrary to this, our cDNA sequencing suggests PfALAD is at least 1.8 kb in size, with Northern analysis (Fig. 6) indicating it may be up to 2.6 kb.

To begin analyzing how such unusual genes are transcribed, we performed a Northern blot. We used three separate probes for the blot. A probe specific to PfSPP revealed a transcript of around 6.6 kb and a probe specific to PfALAD revealed a clear band at ~ 2.6 kb, whereas a probe to the common exons of PfSPP and PfALAD revealed two bands of ~ 6.6 and 2.6 kb in size (Fig. 6). The two bands identified with the common probe correspond to the sizes of the single bands observed for the two gene-specific probes. This confirms the cDNA analysis showing that the two 5' exons of the two transcripts are identical and suggests that alternative splicing after these two exons produces the separate PfSPP and PfALAD transcripts. Significantly, the PfSPP band (6.6 kb) is considerably larger in size than that of PfALAD (2.6 kb).

DISCUSSION

This paper describes processing of the leader sequence of an apicoplast-targeted protein in P. falciparum and presents a candidate for the enzyme that performs this cleavage reaction. Together, the pulse-chase experiment (Fig. 1) and the N-terminal sequencing (Fig. 2) provide direct evidence that transit peptide regions of apicoplast-targeted proteins are post-translationally removed. The pulse-chase experiment also provides information about the kinetics of the import process. The time taken for the unprocessed band to disappear over the chase corresponds to the time it takes for proteins to transport from the ER into the apicoplast, where the cleavage presumably takes place. Because a small amount of processed protein was already visible after the 45-min pulse (Fig. 1), we can infer that the minimum time from synthesis to cleavage is less than 45 min. We attempted to determine the minimum time for processing, but shorter pulses did not provide enough labeling to visualize the proteins. Nevertheless, the relative amounts of precursor versus processed protein at the various chase inter-

 $^{^{3}\,\}mathrm{B.}$ Foth, S. Ralph, C. Tonkin, and G. McFadden, unpublished information.

Leader Cleavage in Malarial Plastids

	* ^*	<u> </u>
ArabidopsisSPP PeaSPP	H VGS I DE E DE QGIAHM I E H VAF LGSKK RE KLL GT H VGS I DE E D DE QGIAHM I E H VAF LGSKK RE KLL GT	GARS-N GARS-N
PfalciparumSPP BuccliiSPP	NTGSVNEKKNQQGISHLCEHVSYMGSKNRKNIVDK	NIRT-N
TparvaSPP	CSGSADE KDNQ RGIAHLCEHVTYMGSKKRTRLCNY	SVKT-N
BorreliaZnProte	NVGSLNEEDNERGIAHYLEHMAFNGTKDYPGNSIVDVLKKFGMQF HTGSLOBEDNELGVAHEVEHMMENGTKTWPGNKVLETFESMGLEF	GADI-N
SpombeMPPB	- AGSRAE TAKN NGAAHFLEHLAFKGTKN RSQKALELEFGNT	GAHL-N
ArabidopsisMPPB PatNardilysin	- AGSRFE SDETNGTAHFLEHMIFKGTDR RTVRALEEEIEDI GVGSFAD PDDLPGLAHFLEHMVEMGSLK VPDENGEDAFLKKH	GGHL-N CGSD-N
RatMPPB	- AGSRYE NEKN NGTAHFLEHMAFKGTKK RSQLDLELEIENM	GAHL-N
PlasmodiumMPPB Synechocystis2	- SGSKYE N KKN NGVAHFLEHMIFKGTKK RN RIQLEKEIENM - AGATAE PDAWPGVAHLIEHMIFKGTKR VPPGAFDOVIEVN	GAHL-N CGMA-N
Synechocytsis1	RVGSRWEGDEINGTAHFLEHMVFKGTPRLAMGEFERAIESR	GAGT-N
HumanInsulysin EimeriaSDP	H IGSLSDPPNIAGLSHFCEHMLFLGTKKYPKENEYSQFLSEH NTGSLYDPODVPGLAHFLEHMLFLGWSKYPEPESYDSFLTE	AGSS-N SGGANN
EcoliPitrilysin	PVGSLED PEAYQGLAHYLEHMSLMGSKK YPQADSLAEYLKMH	GGSH-N
ArabidopsisMPPA PlasmodiumMPPA	- CGSIYDTPOFREATELDERMASKSWLNRSHFRLVREIEAI - CGSRYDEINDOVNEOGMSVMDDNMACHSWAHLSHLRTIKSLEKI	GATV-S
PlasFalcilysin	AFYVKTLTHSGKGIPHILEHSVLSGSKNYNYKNSIGLLEKGTL	HTHL-N
consensus		••••
ArabidoneieSPP		
PeaSPP	AYTDFHHTVFHIHSPTSTKDSD	
PfalciparumSPP ProcliiSPP	A YTDFHHIVFYISVSLNNEIYKENFYSDFKCDQFIKNIREEDIEDYDIY	TVDEFN
TparvaSPP	AFTDYNQTVFFVKTDSELGCNT	
BorreliaZnProte	A A T S F D F T Y Y R L D L S D G N N K D E	
SpombeMPPB	A Y T S REQ TV YYA HAP QNA VPH	
ArabidopsisMPPB	A Y T S R E Q T T Y Y A K V L D S N V N Q	
RatNardilysin RatMPPB	A ST D C E R TV F O F D V Q R K Y F K E	
PlasmodiumMPPB	AYTAREQTGYYCKCFKNDIKW	
Synechocystis 2 Synechocytsis 1	MANSHDYAH FML TTAADYLPR	
HumanInsulysin	AFTSGEHTNYYFDVSHEHLEG	
EimeriaSDP EcoliPitrilysin	A X T D E E K T V F F N K V T D S S F E E	
ArabidopsisMPPA	ASASREQMGYTI DALKTYVPE	
PlasmodiumMPPA PlasFalcilysin	CNAFREHMVYSCECLKEYLPI	
consensus		
	A *	
ArabidopsisSPP	DLFPSVLDAENEIAF-HPKELSSRVEKERRAILSELOMMNTIEYRV	DCQLLQ
ArabidopsisSPP PeaSPP PfalciparumSPP	A * D LFP SVLDALNE IAF - HP KFL S SR VE KERRAIL SELQMMNT I EY RV D LLP SVLDALNE ITF - HP NFL A SR I E KERRAIL SELQMMNT I EY RV Y KHAIL SOCIDT MVE VLK GET OFN KER I T KEK KAIF SEYSI I NN I EY KM	DCQLLQ DCQLLQ NSDIIK
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP	A * D LFP SVLDA INE IAF - HP KFL S SR VE KERRAILSEL QMMNT I E Y RV D LLP SVLDA INE ITF - HP NFL A SR I E KERRAILSEL QMMNT I E Y RV Y KHA I LS QC IDT MVE VLK GET QFN KER I T KEK KAIF SE Y SI INN I E Y KM Y KHS I LS QC LD V MVD VLS GRD QFN KER I T KEK KAIF SE Y SI INT VE Y KI	DCQLLQ DCQLLQ NSDIIK NSDIIK
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte	A D L F P S V L D A L NE I A F - H P K F L S S R VE KER RAIL SE L QMMNT I E Y R V D L L P S V L D A L NE I T F - H P N F L A S R I E K B R RAIL SE L QMMNT I E Y R V Y K H A I L S Q C I D T M VE V L K G E T Q F N K E R I T K E K KAI F S E Y S I I N N I E Y K M Y K H S I L S Q C L D V M V D V L S G R D Q F N K E R I T K E K KAI F S E Y S I I N T V E Y K I - K K E K L H K V L D T L D V A P S Q F S Q Y E L D K E R K A V L SE A K I I T T A E Y K I D E S I N I M W A S - O I S F M K E I D L D R N I T I E K K L G E T Y P G R I - K C E K L B K L G E T Y P G R I I D E S I N I B R W A S - O I S F M K E E I D L D R N I T I E K K L G E T Y P G R I I D E S I N I B R M W A S - O I S F M K E I D L D R N I T I E K K L G E T Y P G R I I D E S I N I B R M W A S - O I S F M K E I D L D R N I T I E K K L G E T Y P G R I I D E S I N I B R M W A S - O I S F M K E I D L D R N I T I E K K L G E T Y P G R I	DCQLLQ DCQLLQ NSDIIK NSDIIK NCNTVK YEKMDK
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease	A * D LFP SV LD A INE IAF - HP KFL SSR VE KER RAILSE L QMMNT I EY RV D LLP SV LD A INE ITF - HP NFL ASR IE KER RAILSE L QMMNT I EY RV Y KHA I LS QC IDT MVE VLK GET QFN KER ITKEK KAIF SEYSI INN I EY KM Y KHS I LS QC LD V MVD VLS GRD QFN KER ITKEK KAIF SEYSI INN I EY KM - KKE KLH KV LD T LD VAPS - QFS QYE LD KER KAVLSE AKI ITT EY KK ID ES INI UR WAS - QI SFM KEE ID LER NI I IEE KKL GET YPG RI LQ QV MAIFSE WSN - AATEE KLE VD AER GVITEE WRAHQD AKWRT	DCQLLQ DCQLLQ NSDIIK NSDIIK NCNTVK YEKMDK SQARRP
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB	A * D LFP SV LD A INE IAF - HP KFL SSR VE KER RAILSE L QMMNT I EYRV D LLP SV LD A INE ITF - HP NFL ASR IE KERRAILSE L QMMNT I EYRV Y KHA I LS QC IDT MVE VLK GET QFN KER ITKEK KAIFSE YSI INN I EYRM Y KHS I LS QC LD V MVD VLS GRD QFN KER ITKEK KAIFSE YSI INT VEYKI - KKE KLHK VLD T LD VAPS - QFS QYE LD KER KAVLSE AKI ITT AEYYK I D ESINI IRN WAS - QI SFM KEE ID LERN II I EEKKL GET YPGRI LQ QV MAIFSE WSN - AATFE KLE VD AER GVITEE WRAHQD AKWRT AVAVIAD ILT - NS KIE EOR IN RER DVIL LE QEE VDK MAD EV	DCQLLQ DCQLLQ NSDIIK NSDIIK NCNTVK YEKMDK SQARRP VLDHLH
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin	A * D LFP SV LD A INE IAF - HP KFL SSR VE KER RAILSE L QMMNT I EYRV D LLP SV LD A INE ITF - HP NFL ASR IE KERRAILSE L QMMNT I EYRV Y KHA I LS QC IDT MVE VLK GET QFN KER ITKEK KAIFSE YSI INN I EYRM Y KHS I LS QC LD V MVD VLS GRD QFN KER ITKEK KAIFSE YSI INT VEYKI - KKE KLHK VLD T LD VAPS - QFS QYE LD KER KAVL SE AKI I TTAEYYK ID ESINI RN WAS - QI SFM KEE ID LERN I I IEEKKL GET YPGRI LQ QV MAIFSE WSN - AATEE KLE VD AER GVITEE WRAHQD AKWRT AVAV IAD ILT - NS KIE E QR IN RERD VIL RE MQE EVD KMAD EV ALD V MAD ILQ - NS KEE EQ RIN RERD VIL RE MQE VEG QTD EV ALD R WAQFFI - HP LMIRDA ID REV EAVD SE Y QL ARP SDANR	DCQLLQ DCQLLQ NSDIIK NSDIIK NCNTVK YEKMDK SQARRP VFDHLH VLDHLH KEMLFG
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB	A * DLFPSVLDA INE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV DLLPSVLDA INE ITF-HPNFLASR IEKERRAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSI INN IEYRM YKHSILSQCLDVMVDVLSGRDQFNKER ITKEKKAIFSEYSI INTVEYKI -KKEKLHKVLDTULDVAPSQFSQYELDKER KAVLSEAKI ITTAEYYK IDESINI RNWAS-QISFMKEE IDLERNI IEEKKLGETYPGRI LQQVMAIFSEWSN-AATEKLEVDAER GVITEEWRAHQDAKWRT AVAVIADILT-NSSISASAVERER QVILRE QEEVDKMADEV ALDVHADILC-NSKEEQRINRERDVILRE MQEVEGQTDEV AVEIIADILQ-NSKEEQRINREVLREMQEVETNLQEV	DCQLLQ DCQLLQ NSDIIK NSDIIK NCNTVK YEKMDK SQARRP VFDHLH VFDHLH KEMLFG VFDYLH
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2	A * DLFPSVLDAINE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSI INN IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSI INTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKER ITKEKKAIFSEYSI INTVEYKI -KKEKLHKVLDTLDVAPSQFSQYELDKER KAVLSEAKI ITTAEYYK LQQVMAIFSEWSN-AATEKLEVDAER GVITEEWRAHQDAKWRT AVAV AADILT-NSSISASAVERER QVILEEWRAHQDAKWRT 	DCQLLQ DCQLLQ NSDIIK NSDIIK NCNTVK YEKMDK SQARRP VFDHLH VLDHLH KEMLFG GFQALC
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocytsis1 HumanInsulysin	A * DLFPSVLDA INE IAF-HPKFLSSRVEKERRAILSELQMMNT IEYRV DLLPSVLDA INE ITF-HPNFLASRIEKERRAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINN IEYRM YKHSILSQCLDVMVDVLSGRDQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTULDVAPS-QFSQVELDKERKAVLSEAKIITTAEYYK IDESINILRNWAS-QISFMKEEIDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATEEKLEVDAERGVITEEWRAHQDAKWRT AVAV ADILT-NSSISASAVERERQVITEEWRAHQDAKWRT ALDVHADILLONSKEEQRINRERDVILREMQEVEKKADEV 	DCQLLQ NSDIIK NSDIIK NCNTVK SQARRP VFDHLH KEMLFG VFDYLH GFQALC IFQQVV LFOLE
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocytsis1 HumanInsulysin EimeriaSDP	A * DLFPSVLDA INE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSIINN IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKER ITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTLLDVAPS-QISFMKEEIDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATEEKLEVDAERGVITEEWRAHQDAKWRT AVAVVADILT-NSSISASAVERER QVIIEEWRAHQDAKWRT 	DCQLLQ NSDIIK NSDIIK NCNTVK SQARRP VFDHLH KEMLFG VFDYLH GFQALC LFQLVK LFQLVK LFQLVK
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocytsis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA	A * DLFPSVLDA INE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV DLLPSVLDA INE ITF-HPNFLASRIEKERRAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTULDVAPS-QISFMKERITKEKKAIFSEYSIINTVEYKI LQQVMAIFSEWSN-AATEEKLEVDAERGVITEEWRAHQDAKWRT AVAV AD ILT-NSSISASAVERERQVITEEWRAHQDAKWRT ALDVHADILQ-NSKFEEQRINRERDVILREMQEVEKGQTDEV ALDVHADILQ-NSKFEEQRINRERDVILREMQEVEKCKDEV 	DCQLLQ NSDIIK NSDIIK NCNTVK SQARRP VFDHLH KEMLFG VFDYLH GFQALC IFQLV LFQLV LFQLV LFQLV LFQLV LFQLV LFQLV
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocytsis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA	A * DLFPSVLDA INE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV DLLPSVLDA INE ITF-HPNFLASRIEKERRAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTULDVAPS-QISFMKEEIDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATEKLVDAERGVITEEWRAHQDAKWRT AVAV AD ILT-NSSISASAVERER QVILEEWRAHQDAKWRT 	DCQLLQ NSDIIK NSDIIK NCNTVK SQARR VFDHLH KEMLFG VFDYLH GFQALC IFQQVV LFQLV LFQLV LFQLV LFQLV LLEAV SA LLEAV
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsiSMPPA PlasmodiumMPPA PlasFalcilysin consensus	A * DLFPSVLDA INE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV DLLPSVLDA INE ITF-HPNFLASRIEKERRAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTLLDVAPS-QFSQVELDKERKAVLSEAKIITAEYYK IDESINIERNWAS-QISFMKEEIDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATEEKLEVDAERGVITEEWRAHQDAKWRT AVAV JADILT-NSSISASAVERERQVITEEWRAHQDAKWRT AVAV JADILLONSKEEQRINRERDVLREMQEVEGQTDEV ALDVHADILQ-NSKEEQRINRERDVLREMQEVEGQTDEV 	DCQLLQ NSDIIK NSDIIK NCNTVK SQARHUK VFDHLH KEMLFG GFQALC VFDYLH GFQALC LFQLV KLH GFQALC LFQLV KLH IFDKLH IFDKLH IFQV SIR AQVSA
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus	A * DLFPSVLDA INE IAF-HPKFLSSRVEKER RAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKRAILSELQMMNT IEYRV YKHAILSQCIDTMVEVLKGETQFNKER ITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKER ITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTLLDVAPSQFSQYELDKER KAVLSEAKIITTAEYYK 	DCQLLQ NSDIIK NSDIIK VEKMDK VEKMDK VLDHLH KEMLFG ULDHLH IFDXLH IFQLV IFQLV XLD IFQLV XLD IFQLV XLD IFQLV XLD IFQLV X X X X X X X X X X X X X X X X X X X
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus	A * DLFPSVLDA NE IAF-HPKELSSRVEKERRAILSELQMMNTIEYRV DLLPSVLDA NE ITF-HPNELASRIEKERRAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVDTLLDVAPSQFSQVELDKERKAVLSEAKIITTAEYYK IDESINIERNWAS-QISFMKEEIDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATEKLEVDAERGVITEEWRAHQDAKWRT ALDVMADILT-NSSISASVEREQVITEEWRAHQDAKWRT 	DCQLLQ NSDIIK NSDIIK VEKMDK VEKMDK VLDHLH KEMLFG UFQLLH IFQLLH IFQLLH IFQLLH IFQLLH IFQLLH IFQLLH IFQLLH IFQLL IFQL IFQL IFQL IFQL IFQL IFQL IFQ
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PeaSPP	A * DLFPSVLDA NE IAF-HPKELSSRVEKER RAILSELQMMNTIEYRV DLLPSVLDA NE ITF-HPNELASRIEKER RAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTLLDVAPSQFSQYELDKERKAVLSBAKIITTAEYYK IDESINIERNWAS-QISFMKEEIDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATEEKLEVDAERGVITEEWRAHQDAKWRT ALDVMADILT-NSSISASVEREQVITEEWRAHQDAKWRT 	DCQLLQ NSDIIIK NSDIIK VEKMUSUIIK VEKMUSU VLDHLH VLDHLHG UFQUVK MAQVSIRA MAQVSIRA MAQVSIRA MAQVSIRA MAQVSIRA VUCC
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PeaSPP PfalciparumSPP	A * DLFPSVLDAENE IAF-HPKFLSSRVEKER RAILSE LOMMNT IEYRV VKHAILSQCIDTMVEVLKGET OFNKER ITKEKRAILSE LOMMNT IEYRV YKHAILSQCIDTMVEVLKGET OFNKER ITKEKKAIFSE YSIINTVEYKI KKSILSQCLDVMVDVLSGRDOFNKER ITKEKKAIFSE YSIINTVEYKI -KKEKLHKVLDTLLDVAPSQFSQYELDKBRKAVLSBAKIITTAEYYK IDESINIERNWAS-QISFMKEE IDLERNIIIEEKKLGETYPGRI LQQVMAIFSEWSN-AATFEKLEVDAERGVITEEWRAHQDAKWRT ALDVHADILT-NSSISASVEREQVITEEWRAHQDAKWRT ALDVHADILQ-NSKEEEQRINRERDVILREMQEVEGQTDEV 	DCQLLQ NSDIIK NSDIIK VEKMUSN VEKMUS VIDHLH KEDLFQLU VIDHLH IFDKLLC VIDKLH IFQLU VIDKL VIDKL VIDKL VIDKL VIDKL VIDKL VIDKL VIDKL VIDKL VIDK VIDKL VIDK VIDK VIDK VIDK VIDK VIDK VIDK VIDK
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PeaSPP PfalciparumSPP TparvaSPP	A * DLFPSVLDA NE IAF-HPKELSSRVEKER RAILSELQMMNTIEYRV DLLPSVLDA NE ITF-HPNELASRIEKER RAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI 	DCQLLQ CQLLQ NSDIIK NSDIIK VEKMUSII VEKMUSI VEKMUSI VECMUSI VECMUSI VECMUSI NACON VECMUSI
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte	A * DLFPSVLDA NE IAF-HPKELSSRVEKER RAILSELQMMNTIEYRV VKHAILSQCIDTMVEVLKGETQFNKERITKEKRAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI 	DCQLLQ NSDIIIK NSDIIK VEKMUSUIIK VEKMUSU VEKMUSU VECMUSIA VECMUSIA NAQUE VECMUSIA VECMUSIA VECMUSIA VECMUSIA VECMUSIA VECMUSIA VECMUSIA VECMUSIA VECMUSIA
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin Consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease	A * DLFPSVLDA NE IAF-HPKELSSRVEKER RAILSELQMMNTIEYRV VKHAILSQCIDTMVEVLKGETQFNKERITKEKRAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVLDTLLDVAPSQFSQYELDKERKAVLSBAKIITTAEYYK IDESINIERNWAS-QISFMKEEIDLERNIFIEEWRAHQDAKWRT LQQVMAIFSEWSN-AATFEKLEVDAERGVITEEWRAHQDAKWRT ALDVMADILT-NSSISASVEREQVITEEWRAHQDAKWRT 	DCQUIIK NSDIIIK QQDIIIK NSDIIIK VEKARLLH VEKARLLH VECARLLH VEDUKLLCV VEDUKLLCV VEDUK
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasmodiumMPPA PlasFalcilysin Consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin	A * DLFPSVLDA NE IAF-HPKELSSRVEKER RAILSELQMMNTIEYRV VKHAILSQCIDTMVEVLKGETQFNKERITKEKRAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI YKHSILSQCLDVMVDVLSGRDQFNKERITKEKKAIFSEYSIINTVEYKI -KKEKLHKVUDTLLDVAPSQFSQYELDKBRKAVLSEAKIITTAEYYK IDESINIERNWAS-QISFMKEEIDLERNIFIEEWRAHQDAKWRT LQQVMAIFSEWSN-AATFEKLEVDAERGVITEEWRAHQDAKWRT ALDVHADILT-NSSISASVEREQVITEEWRAHQDAKWRT 	DCQUIIK NSDIIIK NSDIIIK VEKARLH VEKARLH VEKARLH VECKARLH VECKARLH VECKARLH VECKARLH VECKARL VE
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocytsis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasmodiumMPPA PlasFalcilysin Consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB	A * DLFPSVLDA NE IAF-HPKELSSRVEKER RAILSELQMMNTIEYRV VKHAILSQCIDTMVEVLKGETQFNKERITKEKRAILSELQMMNTIEYRV YKHAILSQCIDTMVEVLKGETQFNKERITKEKKAIFSEYSIINTVEYKI 	DCQUIIK NSDIIIK QQDIIIK NSDIIIK VEKARLH VEKARLH VEKARLH VECKARL VECKARLH VECKARLH VECKARLH VECKARLH VECKARLH VECKARL VECKARLH VECKARL VECKARL VECKARL VECKARL VECKARL VECKARL VECKARL VECKARLH VECKARLH VECKARL VECKARLH VECKARL VECKARLH VECKARLH VECKARLH VECKARLA VECKARLA
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin Consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB ArabidopsisMPPB BatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2	<pre>A</pre>	CQLLQQ NSDIIIK NSDIIIK VEKMUSU VVLDHLHGG VVLDHLHGG VVLDHLGG VVLDLHCVVX VFDVLDLLCVVX VFDVLDLLCVVX VVVVVCD LFQLVVVCCD LYVVCCD VVVVCCD VVVVCCD VVVVCCD VVVVCCD VVVVCCD VVVVCCD
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin Consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB BatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1	<pre>A</pre>	DCQLIIK VCQLIIK NSDIIIK VCQLIIK VCQLIIK VCCKARLH VCCKARLH VCCKARLH VCCKARLH VCCKARLH VCCKARLH VCCC VCC VCC VCC VCC VCC VCC VCC VCC V
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 Human Insulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB ArabidopsisMPPB PlasmodiumMPB BatNardilysin RatMPPB PlasmodiumMPPB Synechocytsis1 HumanInsulysin	A * DLLPSVLDAINE IAF-HPKFLSSR VEKERRAILSELQMMMNTIEYRV VKHAILSQCIDTMVEVLKGET OFNKERITKEKERAILSELQMMNTIEYRV VKHAILSQCIDTMVEVLKGET OFNKERITKEKAAIFSEYSIINNIEYKM YKHSILSQCIDTMVEVLKGET OFNKERITKEKAAIFSEYSIINNIEYKM YKHSILSQCIDTMVEVLKGET OFNKERITKEKAAIFSEYSIINNIEYKM YKHSULSQCIDTMVEVLKGET OFNKERITKEKAAIFSEYSIINNIEYKM YKHSULSQCIDTMVEVLKGET OFNKERITKEKAAIFSEYSIINNIEYKM 	DCQLLQQ NNSDIIK NYEKMULH VFDVLDHLHGG VFDVLDHLHGG VFDVLDHLHGG VFDVLDHLFGU VFDVLD VFDVLD VFDVLLCV VFDVLD VFDVLLCV VFDVLCC VFVVCCD VVCCD VVCD V
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis1 Human Insulysin EimeriaSDP EcoliPitrilysin Consensus ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB ArabidopsisMPPB BarabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocytsis1 HumanInsulysin EimeriaSDP EcoliPitrilysin	<pre>A</pre>	DCQLLQQ NSDIIK NSDIIK VEKMUSDIIK VEKMUSDIIK VFDHLHGG VFDULHLGG VFDULHLGG VFDULHLGG VFDULHCV VFDULHLG VFDULHCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULHLCV VFDULLCV VFDUCCUNCCUNC VFUCCUNCUNCUNCUNCUNCUNCUNCUNCUNCUNCUNCUNCU
ArabidopsisSPP PeaSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocystis2 Synechocystis1 Human Insulysin EimeriaSDP EcoliPitrilysin Consensus ArabidopsisMPPA PlasmodiumMPPA PlasFalcilysin consensus ArabidopsisSPP PfalciparumSPP PyoeliiSPP TparvaSPP BorreliaZnProte EcoliZnProtease SpombeMPPB ArabidopsisMPPB RatNardilysin RatMPPB PlasmodiumMPPB Synechocystis2 Synechocytsis1 HumanInsulysin EimeriaSDP EcoliPitrilysin ArabidopsisMPPB	A * 	CQLLQQ LLQQ CQLIIK NSDIIK NSDIIK VEXMUSS VFDHLHGG VFDVLDHLGG VFDVLDHLGG VFDVLDHLGG VFDVLD VFDVLC VFDVLC VFDVLC VFDVLC VFDVLC VFDVLC VFDVLC VFVVC VFVVC VCC VFVVC VCC VFVVC VCC VC
Ar abi dop sis SPP Pe aSPP Pf al c ip ar um SPP Py oel ii SPP Tp ar va SPP Borrelia ZnProte Ecoli ZnProtease Sp omb eMPPB Ar abi dop sis MPPB RatNardi lysin RatMPPB Pl asmodi umMPPB Synechocystis 2 Synechocystis 1 Human Insulysin Eimeria SDP Ecoli Pitril ysin Ar abi dop sis MPPA Pl asmodi umMPPA Pl asmodi umMPPA Pl asFalci lysin consensus Ar abi dop sis SPP Pe aSPP Pf al ci par um SPP Py oel ii SPP Tp ar va SPP Borrelia ZnProte Ecoli ZnProtease Sp omb eMPPB Ar abi dop sis MPPB RatNardi lysin RatMPPB Pl asmodi umMPPB Synechocystis 2 Synechocytsis 1 Human Insulysin Eimeri aSDP Ecoli Pitril ysin Ar abi dop sis MPPB RatNardi lysin RatMPPB Pl asmodi umMPPB Synechocytsis 1 Human Insulysin Eimeri aSDP Ecoli Pitril ysin Ar abi dop sis MPPA Pl asFalci lysin	A * 	CQLLQQ NSDIIK NSDIIK VEXMUSSON VFDHLHGG VFDVLDHLGG VFDVLDHLGG VFDVLDHLGG VFDVLDHLGG VFDVLD VFDVLD VFDVLC VFDVLC VFDVLC VFDVLC VFDVLC VFDVLC VFDVLC VFVVC VCC VCC VCC VCC VCC VCC VCC VCC

FIG. 4. An alignment surrounding the zinc binding domain of 20 pitrilysin proteins. Zinc binding ligands (*), active site residues (^), and conserved residues (+) are indicated in the alignment following the assignment of Eggleson *et al.* (30). *P. falciparum* SPP contains most of these conserved residues. The exception is the conserved tyrosine (Y), with PfSPP containing a phenylalanine (F) in this position. Of the other *P. falciparum* pitrilysin proteins sequenced in this paper, PfMPPB contains all the conserved residues. PfMPPA lacks many of the conserved residues, including the histidines within the HXXEH motif. A similar observation has been made in other MPP α subunits, which are thought to play a lesser role in the enzymatic activity of MPPs compared with the β subunit (64).



FIG. 5. Schematic structure of *P. falciparum* SPP and ALAD genes. The central panel shows the genomic structure of the two genes, with exons (*E*) shown in *blocks* and introns shown as *thin lines*. The two putative polyadenylation sites are indicated as pA1 and pA2. The cDNAs for PfALAD (*top panel*) and PfSPP (*bottom panel*) share the first two exons (*E1* and *E2*; *green*). PfALAD cDNA has four further exons (*E3–E6*; *red*), with a polyadenylation site (pA1) probably occurring after E6. In PfSPP, these four PfALAD exons are spliced out as an intron, with PfSPP (*bottom panel*), with a putative polyadenylation site (pA2) occurring after E10. The size of exons and introns are drawn approximately to scale, although it should be noted that the untranslated regions from both cDNA transcripts were not fully sequenced.



FIG. 6. Northern blot on *P. falciparum* RNA using probes specific to SPP and ALAD RNA as well as a probe specific to the shared exons of the SPP and ALAD transcripts. The common probe binds to two transcripts, one corresponding to the size of the message identified with the SPP-specific probe (\sim 6.6 kb) and the other corresponding to the message identified with the ALAD-specific probe (\sim 2.6 kb). The size of these transcripts is somewhat larger than the cDNA that was sequenced, suggesting that parts of the 5'- and/or 3'-untranslated regions remain to be sequenced. The fact that the ALAD transcript is smaller in size than the SPP transcript suggests it may contain a different polyadenylation site to the SPP transcript.

vals suggests that the minimum time required for apicoplast targeting is probably around 30-40 min. The ACP leader-GFP fusion protein is under the direction of the calmodulin promoter, a promoter that is likely expressed at high levels (39). In our estimate of processing time, we are assuming that the ACP leader-GFP fusion protein is not expressed at levels where it interferes with the apicoplast import machinery, for instance by saturating apicoplast membrane receptors or the cleavage enzyme (PfSPP). Clearly, analyzing the processing of a native protein rather than an introduced one would address this confounding variable. Attempts to purify native ACP using an anti-ACP antibody produced by Waller *et al.* (5) were unsuccessful.

Assuming this limitation has a minimal effect on the kinetics of apicoplast targeting, how does an apicoplast import time of about 40 min compare with plastid targeting in other organisms? Evidence suggests that less than 10 min is required for plastid targeting in plants and green algae (40), whereas around 40 min is required in *Euglena gracilis* (41). Like apicoplasts, the plastids of *E. gracilis* are derived by secondary endosymbiosis, and nuclear-encoded plastid proteins are targeted via the secretory pathway (41). The greater time it takes for the proteins of these secondarily derived plastids to arrive at their destination compared with plant plastid proteins could, thus, be explained by their trafficking via the somewhat more circuitous secretory pathway.

Cheresh et al. (42) recently described similar analyses of an ACP leader-GFP fusion protein, where pulse label and chase also confirmed that the putative transit peptide-bearing precursor was converted to a processed form. Whereas our ACP leader-GFP fusion protein is under direction of the calmodulin promoter (5) and is maximally expressed at trophozoite and schizont stages of the erythrocyte life cycle (43), Cheresh et al. (42) utilized the histidine-rich protein promoter to drive ACP leader-GFP fusion protein, which results in maximal expression early on (ring stage) in the intraerythrocytic life cycle. Somewhat unexpectedly, Cheresh et al. (42) were able to immunoprecipitate significant amounts of pre-processed protein from outside the parasite (*i.e.* in the supernatant of the saponin lysis). They interpreted this to mean that proteins are targeted to the apicoplast via the parasitophorous vacuole, a compartment that surrounds the outside of the parasite but is still within the host erythrocyte (see review in Ref. 44). We did not attempt to recover ACP leader-GFP fusion protein from the parasitophorous vacuole in our experiments because light microscopic analysis of our parasites has never indicated that GFP is targeted to this compartment. It is possible that the targeting of ACP leader-GFP fusion protein to the parasitophorous vacuole observed by Cheresh et al. (42) is an artifact, with the apicoplast import apparatus not sufficiently well established at ring stage to cope with overexpressed apicoplasttargeted GFP, resulting in the protein exiting the cell via the default secretory pathway. No information about expression of apicoplast-targeted proteins through the erythrocyte stages is yet available, so it is unclear which promoters are suitable when studying apicoplast targeting in transgenic systems. Clearly it will be necessary to track a native apicoplast-targeted protein through the endomembrane system of Plasmodium to resolve the exact route. Our laboratory is currently investigating the role the parasitophorous vacuole may play in apicoplast targeting.

N-terminal sequencing of the processed reporter identified the cleavage motif QVNF \downarrow LNRKN (where \downarrow represents the point of cleavage). This indicates that the transit peptide domain of *P. falciparum* ACP is 24 amino acids, consistent with Western blot analyses (5). Because the fusion protein was designed to contain several amino acids that were almost certainly part of the mature ACP (5), the "context" of the cleavage site should have been maintained, so we believe the cleavage is faithful to the native processing. Clearly, though, it would be desirable to purify and sequence native proteins. A loosely defined cleavage motif has been identified for plant transit peptides (45). This plant motif of (isoleucine/valine)-X-(alanine/ cysteine) \downarrow alanine (where \downarrow represents the cleavage site) is clearly different from the P. falciparum ACP site (Fig. 2), suggesting such a motif is not relevant for apicoplast proteins. A more valid comparison may be to organisms whose plastid is more closely related to the apicoplast. Proteins targeted to the plastids of several other phyla require a bipartite leader similar to that of P. falciparum (see review in Ref. 46), and evidence suggests that the apicoplast shares a common secondary endosymbiotic origin with many of these (47, 48). The cleavage sites of a few heterokont (brown algal and diatom) and cryptomonad proteins have been identified or predicted (49-51), with many containing a methionine residue in the -1 position (50, 51). The cleavage site of one other putative apicoplast-targeted protein, the fatty acid biosynthesis protein FabI, has been determined in *P. falciparum*. This protein has a methionine in the -1position (7), thus conforming to the "methionine at -1" rule of other secondary endosymbionts. However, given that the ACP cleavage site lacks a methionine in this position, not all apicoplast transit peptide cleavage sites conform to this motif.

Fig. 3 is, to our knowledge, the first phylogenetic tree of the pitrilysin family. The mitochondrial-processing peptidases form a clade within the pitrilysin family, with the α - and β -MPP subunits forming groups within this clade. Another subfamily seems to be represented by a clade that includes human insulvsin, rat nardilvsin, Eimeria SDP, and E, coli pitrilysin. The tree includes four pitrilysins from *P. falciparum*, two of which branch with the α - and β -MPP subunits. The food vacuole protein falcilysin (30) seems quite distinct from other members of the group represented on the tree. The remaining P. falciparum pitrilysin branches with plant SPPs, which together form a distinct SPP clade within the family. Indeed, the presence of a putative apicoplast-targeting leader at its N terminus suggests that PfSPP is targeted to the apicoplast. Analysis of the preliminary sequence data from a range of Apicomplexa indicates that this SPP homologue is widespread throughout the group. These putative apicomplexan SPPs are the first non-plant SPPs that have been identified. Although pitrilysin proteins probably have similar catalytic mechanisms (28), the substrates they bind to and cleave are very diverse, ranging from plastid and mitochondrial-targeting peptides, to insulin and hemoglobin fragments. An alignment of PfSPP with other pitrilysins revealed that PfSPP shares most of the conserved residues around the active site (Fig. 4). Given that homology in the group exists around the catalytic site, it might be assumed that regions where the pitrilysins do not match are important in substrate recognition. PfSPP is quite divergent to plant SPPs in these other regions. This may reflect the difference in P. falciparum transit peptides (rich in asparagine and basic residues) compared with plant transit peptides (rich in hydroxylated and basic residues) that must be recognized by their respective SPPs. Our laboratory is currently attempting to determine whether PfSPP actually functions in recognizing and cleaving apicoplast transit peptides.

The Apicomplexa obtained their plastids by secondary endosymbiosis (for review, see Ref. 46). A feature of secondary endosymbioses is that a large proportion of the genes from the nucleus of the endosymbiont are transferred to the host cell nucleus (for review, see Ref. 52), with the genes required for plastid function targeted back to the plastid. An unexpected

finding during this study was that PfSPP shares its two 5' exons (E1 and E2 in Fig. 5), essentially its predicted apicoplasttargeting leader sequence, with a homologue of the heme biosynthesis enzyme ALAD. To our knowledge, this is the first time two plastid-targeted proteins have been shown to share a predicted plastid-targeting leader. One other example of proteins sharing a leader is known. In rice and maize, the ironsulfur subunit of succinate dehydrogenase and the ribosomal protein S14 (*rps14*) share a mitochondrial-targeting sequence (53, 54). In rice, a non-functional copy of *rps14* is present in the mitochondrial genome. Mitochondria, like plastids, arose via an endosymbiotic process, with much of the genetic material of the symbiont transferred to the nucleus of the host (55). It is believed that rice rps14 is an example of how this transfer might occur. The transferred gene inserts into the nuclear genome near a protein already containing a mitochondrialtargeting sequence and uses this existing targeting sequence to get the gene product back into the organelle (53, 55).

Alternative splicing enables PfSPP and PfALAD to share an N-terminal, organelle-targeting leader. This raises the intriguing possibility that one of the two proteins transferred from the endosymbiont nucleus (or perhaps even from the plastid genome) and inserted itself behind an already existing apicoplast leader sequence in the nuclear genome. In the case of PfALAD, this would mean inserting within PfSPP, directly behind the leader, and for PfSPP this would mean inserting behind PfALAD (Fig. 5). With the genomes of several Apicomplexa currently being sequenced, it may be possible in the near future to chart how this situation may have evolved. It is interesting to note that *P. yoelii* has an ALAD homologue inserted within its SPP, whereas this in not the case for the SPP homologue of the more distantly related apicomplexan *T. parva*.

Regardless of whether they transferred from the symbiont nucleus, the plastid genome or from another point within the host cell genome, at some point either PfALAD or PfSPP became associated with the leader sequence of the other. But simply inserting within or behind an existing apicoplast-targeted gene is not in itself sufficient for the protein to become targeted to the apicoplast. First, the transferred gene needs to be spliced in a way for it to be functionally expressed as a protein with a leader. How, then, does this splicing occur?

A Northern blot indicates that the PfALAD transcript is smaller than the PfSPP transcript (Fig. 6). This is a somewhat unexpected result, since we might expect the PfALAD transcript to be larger in size than the PfSPP transcript, given that the PfSPP transcript would have the four PfALAD exons spliced out as an intron, whereas this is not the case for PfALAD. This result is best explained by the PfALAD transcript containing a different polyadenylation site to PfSPP (Fig. 5). Numerous examples are known where alternatively spliced proteins have different polyadenylation sites (for review, see Refs. 56 and 57). In many of these examples, the proximal (i.e. first) polyadenylation site can be regulated. This suggests that the proximal polyadenylation site at the PfSPP/PfALAD locus may be subject to regulation, with transcripts containing the proximal poly(A) site spliced to form PfALAD mRNA and transcripts containing the distal poly(A) site spliced to form PfSPP mRNA. Upon prolonged exposure of the northern blots in Fig. 6, a weak band, larger in size than the SPP band, appears on membranes incubated in both SPP and ALAD specific probes (not shown). It is possible that this band represents transcripts cleaved at the distal polyadenylation site, but where the ALADspecific exons and introns have not been spliced out. It would be interesting to determine whether such transcripts are functional in translation or whether they represent a primary transcript (i.e. heterogeneous nuclear RNA) or indeed whether they

N.-H. (1979) J. Cell Biol. 83, 615-622

- 12. Richter, S., and Lamppa, G. K. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 7463-7468
- 13. Hopkins, J., Fowler, R., Krishna, S., Wilson, I., Mitchell, G., and Bannister, L. (1999) Protist 150, 283–295
- 14. DeRocher, A., Hagen, C. B., Froehlich, J. E., Feagin, J. E., and Parsons, M. (2000) J. Cell Sci. 113, 3969-3977
- Yung, S., Unnasch, T. R., and Lang-Unnasch, N. (2001) Mol. Biochem. Para-sitol. 118, 11–21

- Trager, W., and Jensen, J. B. (1976) Science 193, 673–675
 Laemmli, E. K. (1970) Nature 227, 680–685
 Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 18-53, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- 19. Zeugge, J., Ralph, S., Schmuker, M., McFadden, G. I., and Schneider, G. (2001) Gene 280, 19-26
- 20. Koussevitzky, S., Neeman, E., Sommer, A., Steffens, J. C., and Harel, E. (1998) J. Biol. Chem. 273, 27064–27069
- 21. VanderVere, P. S., Bennett, T. M., Oblong, J. E., and Lamppa, G. K. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 7177-7181
- 22. Kyes, S., Pinches, R., and Newbold, C. (2000) Mol. Biochem. Parasitol. 105, 311-315
- Van den Broeck, G., Timko, M. P., Kausch, A. P., Cashmore, A. R., Van Montagu, M., and Herrera-Estrella, L. (1985) Nature 313, 358–363
 Vollmer, M., Thomsen, N., Wiek, S., and Seeber, F. (2001) J. Biol. Chem. 276,
- 5483-5490
- 25. Smith, S. M., and Ellis, R. J. (1979) Nature 278, 662-664
- 26. Oblong, J., and Lamppa, G. (1992) EMBO J. 11, 4401-4409
- 27. Glaser, E., Sjoling, S., Szigyarto, C., and Eriksson, A. C. (1996) Biochim. Biophys. Acta 1275, 33-37
- 28. Becker, A., and Roth, R. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 3835-3839 29. Reumann, S., and Keegstra, K. (1999) Trends Plant Sci. 4, 302–307
- 30. Eggleson, K. K., Duffin, K. L., and Goldberg, D. E. (1999) J. Biol. Chem. 274, 32411-32417
- 31. Perlman, R. K., Gehm, B. D., Kuo, W. L., and Rosner, M. R. (1993) J. Biol. Chem. 268, 21538-21544
- 32. Kitada, S., Kojima, K., Shimokata, K., Ogishima, T., and Ito, A. (1998) J. Biol. Chem. 273, 32547-32553
- 33. Becker, A. B., and Roth, R. A. (1993) Biochem. J. 292, 137-142
- Fujita, A., Oka, C., Arikawa, Y., Katagai, T., Tonouchi, A., Kuhara, S., and Misumi, Y. (1994) Nature 372, 567–570
- 35. Jelenska, J., Crawford, M. J., Harb, O. S., Zuther, E., Haselkorn, R., Roos, D. S., and Gornicki, P. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 2723-2728
- 36. Sato, S., Tews, I., and Wilson, R. J. M. (2000) Int. J. Parasitol. 30, 427-439
- Smith, A. G. (1988) Biochem. J. 249, 423–428
 Sato, S., and Wilson, R. J. M. (2002) Curr. Genet. 40, 391–398
- Crabb, B. S., Triglia, T., Waterkeyn, J. G., and Cowman, A. F. (1997) Mol. Biochem. Parasitol. 90, 131–144
- 40. Schmidt, R., Gillham, N., and Boynton, J. (1985) Mol. Cell. Biol. 5, 1093-1099
- 41. Sulli, C., and Schwartzbach, S. D. (1996) Plant Cell 8, 45-53
- 42. Cheresh, P., Harrison, T., Fujioka, H., and Haldar, K. (2002) J. Biol. Chem. **277,** 16265–16277
- 43. Orfa Rojas, M., and Wasserman, M. (1995) J. Biochem. (Tokyo) 118, 1118 - 1123
- 44. Lingelbach, K., and Joiner, K. A. (1998) J. Cell Sci. 111, 1467-1475
- 45. Gaval, Y., and von Heijne, G. (1990) FEBS Lett. 261, 455-458 46. van Dooren, G. G., Schwartzbach, S. D., Osafune, T., and McFadden, G. I. (2001) Biochim. Biophys. Acta 1541, 34–53
- 47. Zhang, Z., Green, B., and Cavalier-Smith, T. (2000) J. Mol. Evol. 51, 26-40
- 48. Fast, N. M., Kissinger, J., Roos, D., and Keeling, P. (2001) Mol. Biol. Evol. 18, 418 - 426
- 49. Liaud, M.-F., Brandt, U., Scherzinger, M., and Cerff, R. (1997) J. Mol. Evol. 44, Suppl. 1, 28-37
- 50. Lang, M., Apt, K. E., and Kroth, P. G. (1998) J. Biol. Chem. 273, 30973-30978 51. Deane, J., Fraunholz, M., Su, V., Maier, U.-G., Martin, W., Durnford, D., and

- 53. Kubo, N., Harada, K., Hirai, A., and Kadowaki, K.-I. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 9207-9211
- Figueroa, P., Holuigue, L., Araya, A., and Jordana, X. (2000) Biochem. Bio-phys. Res. Commun. 271, 380–385
- 55. Henze, K., and Martin, W. (2001) Trends Genet. 17, 383-387
- 56. Edwalds-Gilbert, G., Veraldi, K. L., and Milcarek, C. (1997) Nucleic Acids Res. 25, 2547-2561
- 57. Zhao, J., Hyman, L., and Moore, C. (1999) Microbiol. Mol. Biol. Rev. 63, 405 - 445
- Chem. 272, 21839-21846
- (2000) Nat. Med. 6, 898-903
- 60. Kumar, A. M., Schaub, U., Soll, D., and Ujwal, M. L. (1996) Trends Plant Sci. 1, 371–376
- 61. Wilson, C. M., Smith, A. B., and Baylon, R. V. (1996) Mol. Biochem. Parasitol. 79, 135-140
- 1746 1751
- Arch. Biochem. Biophys. 335, 211-218

ALAD catalyzes the second step of the heme biosynthesis pathway, forming porphobilinogen from two molecules of Δ -aminolevulinic acid. Recent experiments have suggested that the majority of ALAD activity in *Plasmodium* spp. is derived from enzymes imported from the host erythrocyte (58, 59). The northern results presented here suggest that PfALAD is actively transcribed during the intraerythrocytic stage of the life cycle of *P. falciparum*. A recent paper, furthermore, found that PfALAD rescues an E. coli ALAD null mutant (38), suggesting the enzyme is functional in heme biosynthesis. This raises interesting questions about the role PfALAD may play in the intraerythocytic cycle. The presence of an apicoplast-targeting leader strongly suggests that PfALAD localizes to the apicoplast. Plant ALAD proteins are similarly targeted to the plastid (37), and PfALAD appears to have a Mg^{2+} binding domain unique to plant ALADs (36). In plants, Δ -aminolevulinic acid is formed via the so-called 5-carbon pathway (for review, see Ref. 60). It is curious, however, that Δ -aminolevulinic acid in *Plasmodium* spp. is formed via the so-called 4-carbon pathway, a single step reaction catalyzed by Δ -aminolevulinic acid synthase (61, 62), an enzyme not found in plants (review in (60)). In other organisms, Δ -aminolevulinic acid synthase is located in the mitochondrion (63). It appears then that enzymes that potentially function in de novo heme synthesis in Plasmodium spp. occur in up to three separate compartments. How the parasite coordinates heme biosynthesis and how such a composite and seemingly complex system evolved is an area that warrants further study.

Acknowledgments-We thank Chris Tonkin for critically reading the manuscript, Kris Ferguson for performing N-terminal sequencing, and Stuart Ralph for sharing ideas. Preliminary sequence data for P. falciparum chromosome 14 was obtained from The Institute for Genomic Research web site (www.tigr.org). Sequencing of chromosome 14 was part of the International Malaria Genome Sequencing Project and was supported by awards from the Burroughs Wellcome Fund and the United States Department of Defense. Sequence data for P. falciparum chromosomes 6, 7, 8, and 9 was obtained from The Sanger Institute web site at www.sanger.ac.uk/Projects/P_falciparum. Sequencing of P. falciparum chromosomes 6, 7, 8, and 9 was accomplished as part of the Malaria Genome Project with support by The Wellcome Trust. Preliminary sequence data from the P. yoelii genome was obtained from The Institute for Genomic Research web site (www.tigr.org). This sequencing program is carried on in collaboration with the Naval Medical Research Center and is supported by the United States Department of Defense. Preliminary sequence data from the T. parva genome was obtained from The Institute for Genomic Research web site at www. tigr.org. Sequencing of the T. parva genome was funded by the International Livestock Research Institute and The Institute for Genomic Research.

REFERENCES

- 1. World Health Organization (1999) The World Health Report 1999, pp. 49-63, World Health Organization, Geneva
- 2. McFadden, G. I., and Roos, D. S. (1999) Trends Microbiol. 6, 328-333
- 3. Ralph, S. A., D'Ombrain, M. C., and McFadden, G. I. (2001) Drug Resist.
- Update 4, 145-151 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) Proc. Natl. Acad. Sci. U. S. A. 95, 12352–12357
 5. Waller, R. F., Reed, M. B., Cowman, A. F., and McFadden, G. I. (2000) EMBO J. 19, 1794–1802
- 6. Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K.,
- Soldati, D., and Beck, E. (1999) Science 285, 1573-1576
- Surolia, N., and Surolia, A. (2001) Nat. Med. 7, 167–173
 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) J. Mol. Biol. 261, 155–172 9. Soll, J., and Tien, R. (1998) Plant Mol. Biol. 38, 191-207
- 10. Bruce, B. D. (2001) Biochim. Biophys. Acta 1541, 2-21
- 11. Schmidt, G., Devillers-Thiery, A., Desruisseaux, H., Blobel, G., and Chua,

- McFadden, G. (2000) Protist 151, 239-252 52. McFadden, G. I. (1999) J. Eukaryot. Microbiol. 46, 339-346

- 58. Bonday, Z. Q., Taketani, S., Gupta, P. D., and Padmanaban, G. (1997) J. Biol.
- 59. Bonday, Z. Q., Dhanasekaran, S., Rangarajan, P. N., and Padmanaban, G.

- Srivastava, P., and Pandey, V. C. (1998) *Exp. Parasitol.* 88, 60–63
 Yamauchi, K., Hayashi, N., and Kikuchi, G. (1980) *J. Biol. Chem.* 255,
- 64. Striebel, H. M., Rysavy, P., Adamec, J., Spizek, J., and Kalousek, F. (1996)