Evolution of malaria parasite plastid targeting sequences

Christopher J. Tonkin*†, Bernardo J. Foth*‡, Stuart A. Ralph*§, Nicole Struck*¶, Alan F. Cowman†, and Geoffrey I. McFadden*∥

*School of Botany, University of Melbourne, Melbourne, Victoria 3010, Australia; and †Division of Infection and Immunity, The Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria 3050, Australia

The transfer of genes from an endosymbiont to its host typically requires acquisition of targeting signals by the gene product to ensure its return to the endosymbiont for function. Many hundreds of plastid-derived genes must have acquired transit peptides for successful relocation to the nucleus. Here, we explore potential evolutionary origins of plastid transit peptides in the malaria parasite *Plasmodium falciparum*. We show that exons of the *P. falciparum* genome could serve as transit peptides after exon shuffling. We further demonstrate that numerous randomized peptides and even whimsical sequences based on English words can also function as transit peptides *in vivo*. Thus, facile acquisition of transit peptides from existing sequence likely expedited endosymbiont integration through intracellular gene transfer.

endosymbiont | gene transfer | protein targeting | transit peptide

Eukaryotic cells are comprised of multiple subcompartments into which different metabolic processes are sequestered. Each compartment has a dedicated suite of proteins responsible for orchestrating specific reactions and functions. Most, if not all, of the proteins in a compartment are encoded by nuclear genes, synthesised in the cytosol, and then targeted to their specific compartment. Targeted proteins typically have a motif or element that is necessary and sufficient to direct the protein to the appropriate compartment. For instance, amphipathic N-terminal helices direct proteins into mitochondria, C-terminal SKL motifs target proteins into peroxisomes, and N-terminal hydrophobic signal peptides mediate translocation of proteins into the secretory system. In plants and algae, N-terminal transit peptides (TPs) direct >1,000 proteins into plastids, but no consensus or common structural elements have yet been identified; how plastids recognise transit peptide-bearing proteins remains largely mysterious (1).

The human malaria parasite *Plasmodium falciparum* contains a nonphotosynthetic plastid (the apicoplast) homologous to chloroplasts of plants (2, 3). The apicoplast harbours indispensable plant-like metabolic pathways that are being inhibited with herbicides and antibacterials to treat malaria (4). More than 400 nucleus-encoded proteins are believed to be targeted to the apicoplast courtesy of bipartite N-terminal extensions comprising a canonical signal peptide followed by a transit peptide equivalent to the transit peptide of plant chloroplast proteins (4–6). After scrutinizing a large collection of these N-terminal extensions from the *P. falciparum* genome, we extracted a simple set of rules and created the bioinformatic tool PlasmoAP that predicts targeting of proteins to the apicoplast from primary sequence (7). Strategic mutagenesis of apicoplast transit peptides demonstrated that a net basic charge and a chaperone-binding site are critical to accurate targeting (7).

Most genes encoding plastid-targeted proteins derive from the endosymbiont and were moved to the nucleus by intracellular gene transfer (8). Transfer of genes is an ongoing process and has led to substantial attrition of the coding capacity of endosymbionts (8). Selection appears to favour gene relocation from plastids to the host nucleus for several reasons. DNA in plastids is thought to be subject to oxidative damage due to photosynthesis so the nucleus is perhaps a safer repository (9). Plastids are also technically haploid and undergo no recombination so moving genes to a sexual milieu in the nucleus may be advantageous (10). Finally, transfer of coding function to the host probably allows better regulation and integration of the endosymbiont (11).

Functional relocation of a plastid gene to the nucleus requires physical movement from one membrane bound compartment to another (8, 12) and acquisition of appropriate expression elements and targeting information to target back to its place of function. Nuclear-encoded plastid proteins require an N-terminal targeting sequence, but how transferred genes acquire this information is unknown. Where do these targeting peptides come from? One possibility was to recycle existing targeting peptides, and two *P. falciparum* proteins share a common exon-encoded targeting motif via alternative mRNA splicing for translocation of the proteins into the plastid (13). However, recycling appears to be infrequent, and most apicoplast-targeted proteins, like most other nuclear-encoded plastid proteins, have unique targeting motifs (14). New targeting sequences can also be generated *de novo*, either by the fortuitous insertion of an unrelated exon (by exon shuffling) or a fragment of random sequence in front of a gene. Both mechanisms are known to have yielded mitochondrion-targeting sequences in plants (15–17). Given that the *P. falciparum* nucleus contains ≈500 genes encoding plastid proteins, each with a unique targeting peptide, we decided to explore mechanisms by which these peptides could have evolved.

In this article, we investigate the possibility that apicoplast-targeting leaders in *P. falciparum* could have originated either by recruitment of unrelated exons through exon shuffling or by the simple acquisition of random pieces of DNA ahead of the gene encoding a protein to be targeted to the plastid. We tested the suitability of *P. falciparum* exons to act as plastid targeting transit peptides both *in silico* and *in vivo*. We also designed and generated various artificial peptides and tested their apicoplast-

EVOLUTION

Table 1. Summary of artificial peptides and their apicoplast targeting capacity

Construct	Peptide sequence	Predicted chaperone binding site (18)	Predicted apicoplast targeting (PlasmoAP)	*In vivo* apicoplast targeting
ACP(s)-GLP-1 (exon 1)	MIMKNKYGVFFSLSKNAIINSNRQ LFPFKKVSKINFSNSADDKNNGVI PEQRTQYSGTNEYKDFEKTEVYQ	+	+	+
ACP(s)-GLP-1 (exon3)	VVNILNSMNVKDYVYIDVMKNNN LREAIKIYSNWPYIPNLYVNNNFIG GYDIISDLYNRGELEKIIK	−	−	−
Artificial TP 1	IKIKKNKKKKNNNIKNNKKKKKKK	−	+	−
Artificial TP 2	KVIKNQKGSDNYLKIRLNHTNSE	+	+	+
Artificial TP 2 MutEE	EVIENQKGSDNYLKIRLNHTNSE	+	−	−
Artificial TP 3	IKIIKNNKISDNSIKIKINKSNSD	+	+	+
Artificial TP 4	KRYNYLQFNKKGKKVIAIKEKINN	+	+	+
Artificial TP 5	LFLLTKNKKYLTIKIESNINTKEK	+	+	+
Artificial TP 6	ELINIKVEHNDLNVKEKTFYAFYC	+	−	−
Artificial TP 6 MutKK	KLINIKVKHNDLNVKEKTFYAFYC	+	+	+
Artificial TP 7	EINKEMIQNIEFFNNNQKNIYHNN	−	−	−
English TP 1	SKINNYSLINKYKINKYTHING	+	+	+
English TP 2	ITWILLNEVERTARGETPLASTID	−	−	−

targeting efficacy. Our results confirm that transit peptides are very simple elements and suggest that they could evolve with relative ease from existing sequence.

Results and Discussion

The bipartite leader of apicoplast-targeted proteins is necessary and sufficient to translocate proteins into the stromal compartment of the relict plastid of apicomplexan parasites (6). If the transit peptide is absent or nonfunctional, then the protein is mis-targeted and accumulates in the default destination—the parasitophorous vacuole, a compartment created by the synthesis of a parasite-derived parasitophorous vacuole membrane that physically separates parasite from host cell (6, 7).

In this study, we tested the ability of various peptide sequences to function as apicoplast-targeting transit peptides in *P. falciparum* when placed behind a functional signal peptide. These tests included both bioinformatic predictions and *in vivo* analysis of transfected parasites, using a fluorescent reporter gene system. The *in silico* predictions were performed by using the apicoplast transit peptide prediction tool PlasmoAP (7), with the PlasmoAP scores given relating directly to the predicted capacity of a peptide to direct green fluorescent protein (GFP) from within the endomembrane system to the apicoplast. In addition, the presence of chaperone (Hsp70)-binding sites, which are thought to also be important for apicoplast targeting (7), was predicted according to Rudiger *et al.* (18). For the *in vivo* analysis, sequences of interest were placed between the signal peptide of acyl carrier protein (ACP) and GFP, followed by transfection into the parasites.

Can Exons Become Transit Peptides? To investigate the feasibility of exon recruitment for transit peptide creation in malaria parasites, we screened the 11,544 sufficiently large exons of the *P. falciparum* genome [see supporting information (SI) *Materials and Methods*] with PlasmoAP (7) and found that 18.3% of these exons have characteristics potentially suitable as transit peptides (SI Table 2). For instance, exon 1 of the gene encoding a glutaredoxin-like protein (GLP-1, gene identifier PFC0205c) is predicted to function as an apicoplast transit peptide (Table 1). Importantly, when placed between a signal peptide and GFP, GLP-1 exon 1 actually does mediate accurate plastid targeting *in vivo* (Fig. 1*A*), corroborating the bioinformatic prediction (Table 1). Conversely, exon 3, which is predicted to have neither the characteristics of a transit peptide nor to contain a chaperone-

binding site, did not target the reporter to the parasite plastid *in vivo* (Fig. 1*B*). Thus, the abundance of exons with transit peptide-like characteristics in the malaria parasite genome, and the frequent occurrence of introns bounding targeting motifs (19, 20) strongly support a role for exon shuffling in the evolution of plastid targeting.

Artificial Transit Peptides. *Manual rearrangements of existing transit peptides.* After having confirmed protein-coding exons as a suitable source for apicoplast-targeting transit peptides, our next question was to ask how likely it was that noncoding or "random" DNA could function as transit peptides. To this end, we designed and synthesised three different types of completely artificial
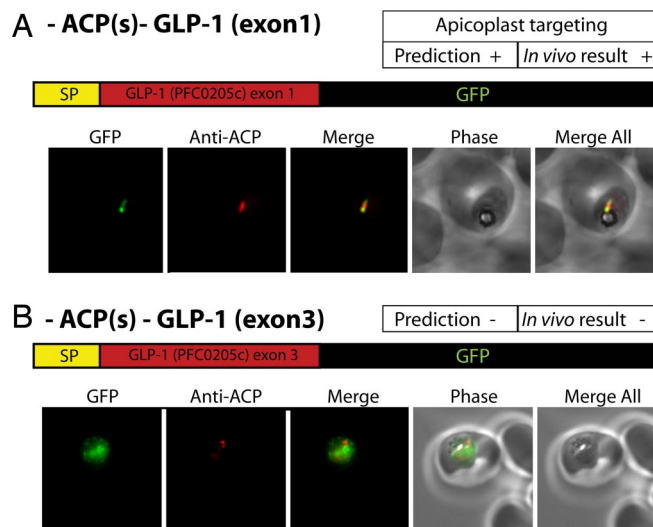

Fig. 1. Efficacy of exons in mediating targeting to the malaria parasite apicoplast. Two exon sequences of the gene encoding the glutaredoxin-like protein 1 (GLP-1), itself a nonapicoplast protein, were incorporated downstream of a signal peptide and upstream of green fluorescent protein (GFP) and expressed in parasites. The images show GFP fluorescence in fixed *P. falciparum* malaria parasites within human red blood cells colabeled with antibodies against the apicoplast marker ACP. For both exons the targeting by PlasmoAP matches the *in vivo* result because exon 1 of GLP-1 mediates targeting of GFP exclusively into the apicoplast (*A*), whereas exon 3 of GLP-1 does not (*B*). For sequences for each of these exons, see Table 1.
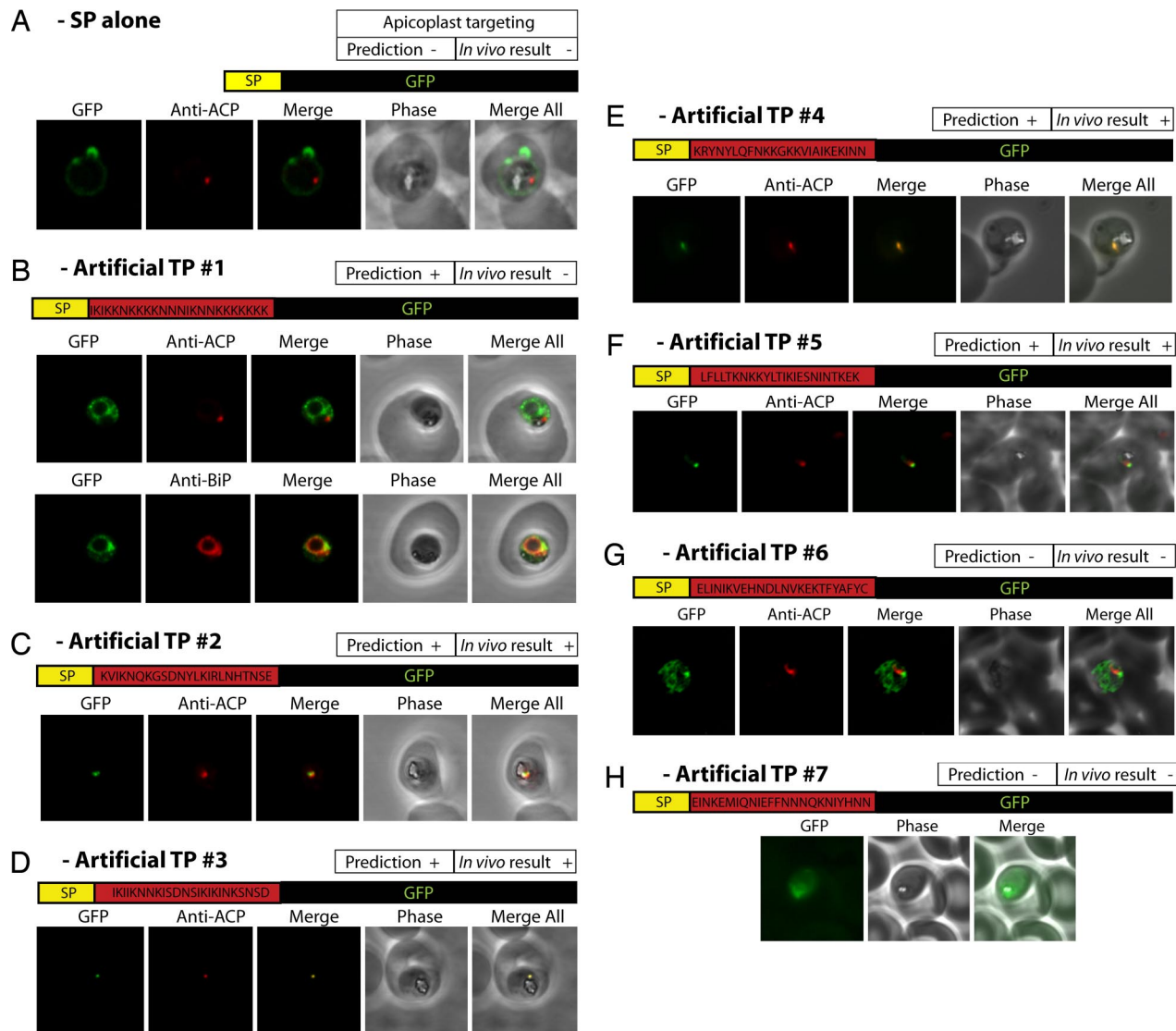
Fig. 2. Experimental assessment of artificial apicoplast transit peptides. The micrographs show the reporter protein GFP in green and the apicoplast-resident protein ACP or the ER-marker BiP in red. (*A*) A control experiment without any putative transit peptide sequence demonstrates that the signal peptide alone directs GFP outside the parasite and into the parasitophorous vacuole but not to the apicoplast. (*B*) Despite a positive targeting prediction by PlasmoAP, Artificial TP 1 fails to direct GFP into the apicoplast. Colocalization with the ER-resident protein BiP shows most of GFP to be present in the ER. (*C–F*) As predicted, Artificial TPs 2–5 mediate targeting of GFP to the apicoplast. (*G* and *H*) As predicted, Artificial TPs 6–7 fail to mediate targeting of GFP to the apicoplast.

peptides 24 aa long (24 mer). First, we created three sequences (Artificial TPs 1–3; Table 1) based on known characteristics of real apicoplast transit peptides. Their design was based on amino acid compositional and positional information gathered from a collection of 76 existing putative apicoplast transit peptides of *P. falciparum* (the "model dataset," available from the authors on request) (7, 14). From these 76 real transit peptides, we took sequence logos, average amino acid contents, and physiochemical properties of amino acids into account (see *SI Materials and Methods*). Two of these artificial peptides (Artificial TPs 2 and 3; Table 1) behaved *in vivo* as predicted by PlasmoAP and targeted the reporter protein to the apicoplast as judged by immunofluorescent colocalization with the apicoplast marker protein ACP (Fig. 2 *C* and *D*). Only artificial transit peptide 1 failed to target the reporter GFP to the apicoplast, despite being predicted by PlasmoAP to be efficacious as a transit peptide (Fig. 2*B*). Rather, it directed the reporter to the ER, because GFP colocalized with the ER-resident protein BiP (Fig. 2*B*). The failure of this peptide could either be due to the lack of a

predicted chaperone binding site or its excessive charge, because it is essentially polylysine (see Table 1).

Randomly scrambled sequences. Based on the success of the first set of artificial peptides and to push the abstraction of artificial transit peptide design one step further, we then computer-generated 10,000 scrambled peptide sequences. For these peptides, the only constraint was that their average amino acid composition be the same as that found in the model dataset of 76 putative apicoplast transit peptides (see *SI Materials and Methods* for details). Screening of the transit peptide potential of these randomized peptides with PlasmoAP revealed that a staggering 29.2% were predicted to be potentially excellent apicoplast transit peptides (see *SI Table 3*). Four of these randomized peptides (two positives and two negatives: Artificial TPs 4 and 5 and Artificial TPs 6 and 7, respectively) were chosen for *in vivo* testing, which fully confirmed the computer predictions (Fig. 2 *E–H* and Table 1). Taken together, these results suggest that numerous random peptides could do duty as transit peptides if they have the appropriate amino acid composition. By
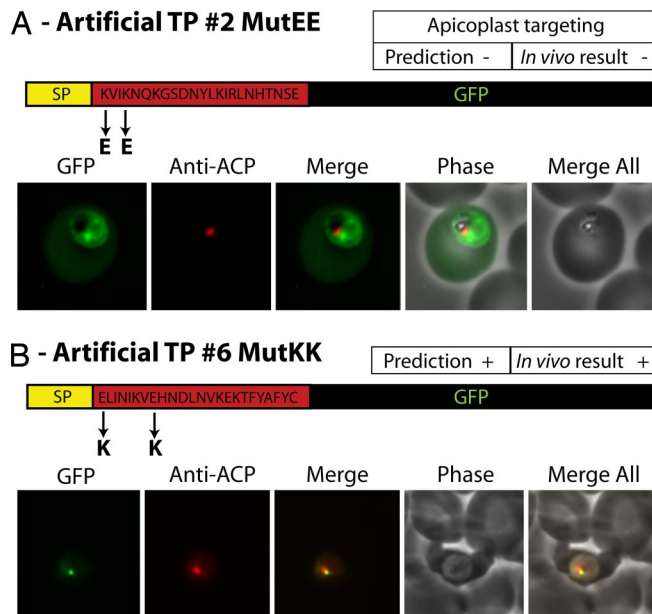
A - Artificial TP #2 MutEE

Apicoplast targeting	
Prediction -	*In vivo* result -

B - Artificial TP #6 MutKK

Prediction +	*In vivo* result +

Fig. 3. Reversal of apicoplast transit peptide functionality of two artificial peptide sequences. Two N-terminal residues in both Artificial TP 2 and Artificial TP 6 were mutagenized to reverse their charge. As for all other *in vivo* experiments described in this study, the peptide sequences were placed downstream of a signal peptide and upstream of green fluorescent protein (GFP), and the images show the reporter protein GFP in green and the apicoplast-resident protein ACP in red. As predicted, switching two charges from positive to negative (two lysine residues, K, mutated to glutamic acid, E) in Artificial TP 2 MutEE abrogates targeting to the apicoplast (*A*), whereas, in the inverse experiment, replacing two negatively charged residues (glutamic acid, E) of Artificial TP 6 MutKK by two positively charged lysines (K) engenders apicoplast targeting of GFP (*B*).

A - English TP #1

Apicoplast targeting	
Prediction +	*In vivo* result +

B - English TP #2

Prediction -	*In vivo* result -

Fig. 4. English phrases as apicoplast transit peptides. (*A*) The English language phrase SKINNYSLINKYKINKYTHING, when expressed as a peptide between a signal peptide and GFP, correctly targets the fluorescent reporter to the apicoplast as seen by colocalisation with the resident apicoplast protein ACP. (*B*) Fusion of ITWILLNEVERTARGETPLASTID peptide in context cannot mediate apicoplast targeting and can be seen mainly localizing to the ER of the parasite.

contrast, 99.9% of truly random 24-mer sequences (in which each of the 20 aa occurred at 5% average frequency) were deemed completely unsuitable as transit peptides by PlasmoAP (− score in SI Table 3). Importantly, we could successfully reverse the targeting behavior of artificial peptides by directed mutagenesis that inverted the charge of the first two positively (Artificial TP 2 MutEE; Fig. 3*A*) or negatively charged amino acids (Artificial TP 6 MutKK; Fig. 3*B*). These results confirm the critical importance of the N-terminal net charge of apicoplast transit peptides (7), even artificial ones.

Genome-wide estimate. Because N-terminal charge and overall amino acid composition are essential determinants of transit peptide functionality, what are the odds that a random piece of *P. falciparum* DNA could produce a functional transit peptide by chance? To address this question, we used PlasmoAP to screen three more sequence collections (see *SI Materials and Methods* for details). We generated 10,000 scrambled peptide sequences with two size classes; one set of 24-mers, and one set of 120-mers to investigate the effect of length. This also was done to explore any effect that the downstream reporter protein sequence had on targeting (however, the GFP sequence has a negative impact on PlasmoAP prediction because of the relative abundance of acidic residues at the GFP N terminus). The 24-mers and the 120-mers in these peptides conformed to the average amino acid composition of the 5,411 predicted proteins of *P. falciparum*. Of the 24-mer peptides, 7.3% received the best PlasmoAP score (SI Table 3). Predictions by PlasmoAP of apicoplast targeting efficacy for the longer 120-mers showed that longer peptides are indeed better transit peptides; 20.3% of the longer peptides received the best score (SI Table 3).

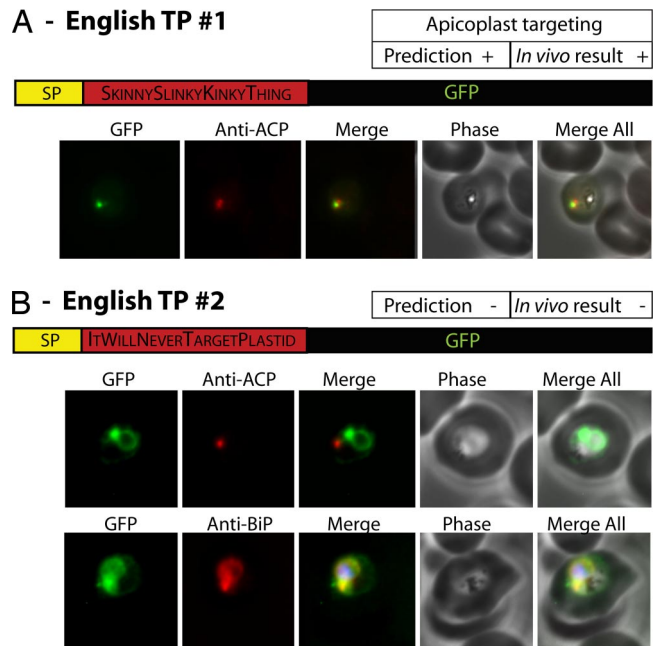Finally, to run an even more realistic simulation, we Plas-moAP-analyzed the complete collection of 101,207 ORFs (ORFs) predicted by PlasmoDB (www.plasmodb.org) to be encoded in the *P. falciparum* genome (all ORFs with a minimum length of 52 aa). This analysis yielded a remarkable 22.3% of sequences receiving the best PlasmoAP score (SI Table 4). Significantly, 94.4% of these PlasmoAP-positive ORFs are also predicted to contain at least one chaperone-/Hsp70-binding site within their first 30 residues (data not shown). In other words, more than one fifth of all ORFs encoded in the *P. falciparum* genome (of at least 52 aa length) are strongly predicted to be able to function as apicoplast-targeting transit peptides if inserted ahead of a gene and downstream of a signal peptide.

English words as transit peptides. To carry the tests of what constitutes a transit peptide to the extremes, we created two peptides with meaning in the English language and introduced these into parasites. The peptide [NH2]SKINNYSLINKYKINKY-THING[COOH] (English TP 1) is predicted to target (Table 1) and does so *in vivo* (Fig. 4*A*). Conversely, [NH2]ITWILLNEVER-TARGETPLASTID[COOH] (English TP 2) is predicted not to target (Table 1) and does not (Fig. 4*B*). Thus, even clearly nonbiological sequences based on capricious phrases and sentences from the English language may mediate plastid targeting *in vivo* as long as they comply with some basic requirements such as an appropriate N-terminal charge, the presence of a chaperone-binding site, and overall hydrophilicity.

Concluding Remarks. We find very strong concordance between the bioinformatic prediction of apicoplast targeting and the *in vivo* behavior of peptide sequences, thus showing that apicoplast transit peptides are truly based on a simple set of parameters. Of the 13 peptide sequences tested—which ranged from genuine *P. falciparum* exons of a nonapicoplast protein to randomized sequences of naturally occurring amino acid composition, and even whimsical sequences comprised of English words—12 peptides behaved *in vivo* as predicted by the bioinformatic tools.

Only one peptide (Artificial TP 1) did not behave as predicted (Fig. 2B). We conclude that virtually any sequence of amino acids with the appropriate positive charge, suitable chaperone-binding site, and general hydrophilicity can act as an apicoplast-targeting transit peptide. Tihs stiuaituon is rmesneiimnct of the way hnumas can stlil uednrtsnad wirtetn lganague eevn if the ltrtees whitin wrods are srcamelbd. (This situation is reminiscent of the way humans can still understand written language even if the letters within words are scrambled). The low complexity of transit peptides has thus probably greatly facilitated their acquisition through exon shuffling and recruitment of "random" sequence (protein coding or noncoding) and expedited intracellular gene transfer during endosymbiosis.

Materials and Methods

Bioinformatic Prediction of Apicoplast Targeting and the Presence of Chaperone Binding Sites. The ability of peptides to be able to function as apicoplast-targeting transit peptides was predicted by using the tool PlasmoAP (7). Because the experimental assessment of apicoplast transit peptide functionality was to be carried out by placing a peptide of interest between a functional signal peptide (that of ACP: MKILLLCIIFLYYVNAF) and GFP, the PlasmoAP algorithm was run without employing the signal peptide search with SignalP, and the protein sequence of GFP (starting PRSKGEELFTGVVPIL-VELD...) was appended to all peptide sequences before the bioinformatic analyses. Potential Hsp70-binding sites were predicted employing the algorithm by Rudiger et al. (18) with a positive prediction being defined as a DnaK binding site prediction score of −5 or smaller. Both types of predictions were carried out by using Perl scripts that are available on request.

Design of Artificial Transit Peptides (TPs). Three different types of artificial transit peptides were created: (*i*) Three sequences (Artificial TPs 1–3) were designed based on the collective features and general organisation—i.e., amino acid compositional and positional information—of 76 highly likely apicoplast transit peptides of *P. falciparum* (7). (*ii*) Another four peptides (Artificial TPs 4–7)—including two sequences designed not to function as apicoplast-targeting transit peptides—were chosen from a computer-generated list of 10,000 randomized 24-mer peptide sequences based on the amino acid composition found in the 76 putative apicoplast transit peptides. (*iii*) Two more artificial transit peptides (English TPs 1 and 2) were designed based on words of the English language that, when understood as amino acid sequences, resulted in the desired predictions regarding apicoplast targeting (according to PlasmoAP) and chaperone binding sites (18). For further details, see *SI Materials and Methods*.

Cloning of Transfection Plasmids. Cloning of exons and artificial transit peptides fused to signal peptides were created by using a variety of methods (see *SI Materials and Methods* for details).

Parasite Transfection. The 3D7 strain of *P. falciparum* was used for all transfections. Parasites were grown in human O+ Red Blood Cells at 4% hematocrit in RPMI medium 1640 supplemented with Albumax (GibcoBRL) to a final concentration of 0.5% and gassed with 5% CO_2 and 0.5% O_2 in N_2 at 37°C as described in ref. 21. Transfection was carried out by electroporation of ring stage parasites (22, 23), using 100 μg of plasmid DNA.

1. Bruce BD (2000) *Trends Cell Biol* 10:440–447.
2. McFadden GI, Reith M, Munholland J, Lang-Unnasch N (1996) *Nature* 381:482.
3. Wilson RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, et al. (1996) *J Mol Biol* 261:155–172.
4. Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz M, Foth BF, Tonkin CJ, Roos DS, McFadden GI (2004) *Nat Rev Microbiol* 2:203–216.
5. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, et al. (2002) *Nature* 419:498–511.
6. Waller RF, Reed MB, Cowman AF, McFadden GI (2000) *EMBO J* 19:1794–1802.
7. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF, McFadden GI (2003) *Science* 299:705–708.
8. Huang CY, Ayliffe MA, Timmis JN (2003) *Nature* 422:72–76.
9. Allen JF (2003) *Philos Trans R Soc Lond B* 358:19–37, discussion 37–18.
10. Blanchard JL, Lynch M (2000) *Trends Genet* 16:315–320.
11. McFadden GI (1999) *Curr Opin Plant Biol* 2:513–519.
12. Stegemann S, Hartmann S, Ruf S, Bock R (2003) *Proc Natl Acad Sci USA* 100:8828–8833.
13. van Dooren GG, Su V, D'Ombrain MC, McFadden GI (2002) *J Biol Chem* 277:23612–23619.
14. Ralph SA, Foth BJ, Hall N, McFadden GI (2004) *Mol Biol Evol* 21:2183–2194.
15. Adams KL, Palmer JD (2003) *Mol Phylogenet Evol* 29:380–395.
16. Long M, de Souza S, Rosenberg C, Gilbert W (1996) *Proc Natl Acad Sci USA* 93:7727–7731.
17. Wischmann C, Schuster W (1995) *FEBS Lett* 374:152–156.
18. Rudiger S, Germeroth L, Schneider-Mergener J, Bukau B (1997) *EMBO J* 16:1501–1507.
19. Schaap D, van Poppel NF, Vermeulen AN (2001) *Mol Biochem Parasitol* 115:119–121.
20. Waller RF, Keeling PJ, Donald RGK, Striepen B, Handman E, Lang-Unnasch N, Cowman AF, Besra GS, Roos DS, McFadden GI (1998) *Proc Natl Acad Sci USA* 95:12352–12357.
21. Trager W, Jensen JB (1976) *Science* 193:673–675.
22. Crabb BS, Cowman AF (1996) *Proc Natl Acad Sci USA* 93:7289–7294.
23. Wu Y, Sifri CD, Lei HH, Su XZ, Wellems TE (1995) *Proc Natl Acad Sci USA* 92:973–977.

EVOLUTION