doi:10.1016/j.jmb.2005.03.030 J. Mol. Biol. (2005) 348, 1015–1024

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Complex Protein Targeting to Dinoflagellate Plastids

Nicola J. Patron†, Ross F. Waller†, John M. Archibald and Patrick J. Keeling*

Canadian Institute for Advanced Research, Program in Evolutionary Biology Department of Botany University of British Columbia 3529-6270 University Blvd. Vancouver, BC, Canada V6T 1Z4

**Corresponding author*

Protein trafficking pathways to plastids are directed by N-terminal targeting peptides. In plants this consists of a relatively simple transit peptide, while in organisms with secondary plastids (which reside within the endomembrane system) a signal peptide is appended to the transit peptide. Despite amino acid compositional differences between organisms, often due to nucleotide biases, the features of plastid targeting sequences are generally consistent within species. Dinoflagellate algae deviate from this trend. We have conducted an expressed sequence tag (EST) survey of the peridinin-plastid containing dinoflagellate *Heterocapsa triquetra* to identify and characterize numerous targeting presequences of plastid proteins encoded in the nucleus. Consistent with targeting systems present in other secondary plastid-containing organisms, these all possess a canonical signal peptide at their N termini, however two major classes of transit peptides occur. Both classes possess a common N-terminal portion of the transit peptide, but one class of transit peptides contains a hydrophobic domain that has been reported to act as a stop-transfer membrane anchor, temporarily arresting protein insertion into the endoplasmic reticulum. A second class of transit peptide lacks this feature. These two classes are represented approximately equally, and for any given protein the class is conserved across all dinoflagellate taxa surveyed to date. This dichotomy suggests that two mechanisms, perhaps even trafficking routes, may direct proteins to dinoflagellate plastids. A four-residue phenylalanine-based motif is also a consistent feature of *H. triquetra* transit peptides, which is an ancient feature predating red algae and galucophytes that was lost in green plastids.

Keywords: dinoflagellate; plastid; protein targeting; transit peptide; red algae

Introduction

The photosynthetic organelles of eukaryotes can be broadly categorized in two groups according to their evolutionary origin, primary plastids and secondary plastids.[1,2] Primary plastids are derived directly from a cyanobacterial endosymbiont, while secondary plastids are derived from the endosymbiotic uptake of a eukaryote that already contained a primary plastid. While all plants and some algae contain primary plastids, secondary plastids are found in dinoflagellates, apicomplexan parasites, heterokont algae, haptophytes, cryptophytes, euglenids and chlorarachniophytes. Amongst these are the major photosynthetic contributors of the oceans (dinoflagellates, diatoms and haptophytes).

The vast majority of proteins in all plastids are encoded by genes in the host nucleus, most of which have been relocated from the plastid genome.[3,4] These proteins are translated by cytosolic ribosomes then targeted across the plastid membranes. In primary plastids, targeting is mediated by an N-terminal serine/threonine-rich extension, known as a transit peptide (TP), that is recognized by translocating-complexes in the two plastidial membranes (Toc and Tic) and subsequently removed by a stromal peptidase.[5,6]

† N.J.P. & R.F.W. contributed equally to this article.

Present address: J. M. Archibald, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada.

Abbreviations used: ER, endoplasmic reticulum; EST, expressed sequence tag; SP, signal peptide; TM, transmembrane; TP, transit peptide.

E-mail address of the corresponding author: pkeeling@interchange.ubc.ca

Protein trafficking to secondary plastids is more complex because proteins must cross one or two additional membranes. The outermost membrane is derived from the host endomembrane system, so proteins targeted to all secondary plastids initiate with a signal peptide (SP) to direct them to the lumen of the endomembrane system, followed by a transit peptide to engage the Tic/Toc apparatus.[4] In heterokonts, haptophytes and cryptophytes the rough endoplasmic reticulum (ER) is physically continuous with the outer plastid membrane, but in groups where this is not the case some form of vesicular transport must exist to direct proteins from the ER to the plastid. Moreover, most secondary plastids are surrounded by an additional membrane between the outer endomembrane and the inner two primary plastid membranes (we will refer to this as the "third membrane"). While the bipartite targeting model can account for the outermost and two inner membranes, the trafficking events that enable proteins to cross the third membrane remain unknown.

Dinoflagellate peridinin-containing plastids and euglenid plastids are unique in having apparently lost the third membrane, and so are surrounded by only three membranes. Paradoxically the presequences that target proteins to these somewhat simpler plastids appear to be more complex: the relatively few plastid-targeting proteins described from both groups are reported to include a third sorting signal. This feature is a hydrophobic domain downstream of the typical hydrophilic serine/threonine-enriched transit peptide. In both cases, this hydrophobic domain has been shown to act as a stop-transfer membrane anchor in canine microsomes[7,8] and preproteins from both groups have been immunolocalized in the Golgi apparatus on route to the plastid (which has not been demonstrated for any other secondary plastids). A model of protein-trafficking to the plastids of both groups has been put forward where proteins are first co-translationally imported into the ER up to the stop-transfer signal, so that the N-terminal region of the transit peptide is in the lumen of the ER while the rest of the protein remains in the cytoplasm. Maintaining this topology, proteins are directed to the Golgi and sorted into vesicles that will fuse with the plastid's outer-most membrane, exposing the transit peptide to the Toc/Tic apparatus, which draws the entire protein across the remaining membranes.[7–11]

Neither the host cells nor plastids of dinoflagellates and euglenids are closely related, so the similarities between their targeting pathways evolved in parallel. Euglenids are related to kinetoplastid parasites such as *Trypanosoma* (which lack a plastid) and have acquired their plastids from a green alga.[12] Dinoflagellates, on the other hand, are most closely related to apicomplexan parasites, and together they are part of the hypothetical supergroup "chromalveolates" (including cryptophytes, haptophytes and heterokonts), which are postulated to share a plastid derived from a single endosymbiosis of a red alga.[13–16] In addition to remodeling plastid membrane number and targeting-presequences, dinoflagellates have relocated the vast majority of plastid-encoded genes to the nucleus[17,18] and reorganized the remaining set of primarily photosystem genes into single-gene mini circles.[19,20] In order to explore the derived features of protein targeting in dinoflagellates, we have assembled a much larger set of full-length plastid-targeted proteins than has previously been available. Owing to the remarkable size of dinoflagellate genomes (3000–215,000 Mb), we have used an expressed sequence tag (EST) survey from the dinoflagellate with the most thoroughly studied peridinin-containing plastid, *Heterocapsa triquetra*, to explore the diversity of plastid-targeted proteins and their presequences. These data reveal even greater complexity of dinoflagellate trafficking to plastids than was previously realized.

Results and Discussion

Identification of plastid-targeted proteins in *H. triquetra*

We generated 6765 *H. triquetra* 5′-end sequenced ESTs that were assembled into 2022 unique clusters. Of these, 63 clusters were identified as likely encoding 52 distinct plastid proteins and 11 predicted isoforms (Table 1). Proteins were aligned to homologs from other organisms, and phylogenetic analyses were inferred to confirm their plastid ancestry (data not shown). The majority of the plastid proteins are involved in the light or dark reactions of photosynthesis, however, functions such as translation and synthesis of fatty acids and isoprenoids were also represented. Based on protein alignments, 32 clusters were predicted to correspond to full-length genes with N-terminal extensions, consistent with the presence of plastid-targeting presequences. A total of 17 other clusters encoded partial presequences, and the remaining 14 clusters were missing regions of the mature protein (Table 1). None of the *H. triquetra* genes identified are known to be represented in the plastid genome of any dinoflagellate, and three genes (ATP synthase subunit C, cytochrome b559, and photosystem II protein L) are plastid-encoded in all other photosynthetic eukaryotes and therefore represent unique gene transfers in dinoflagellates, as has been shown previously.[17,18]

Plastid targeting sequences in *H. triquetra*

Dinoflagellate plastid-targeting presequences have been analyzed from a small number of proteins from several taxa, and only a few have been analyzed in detail.[7,17,18,21,22] The 24 full-length proteins from *H. triquetra* allow a detailed comparative investigation of plastid-targeting motifs (eight isoforms of various proteins were also found, but only one of each was used so as not to bias the

Table 1. *H. triquetra* plastid-targeted proteins predicted from cDNAs

Protein name	Pathway	Accession no.	Full length	Plastid-encoded
Acetolactate synthase (*als*)	Amino acid synthesis	AY826826	5′ Partial	–
Adenylate kinase (*adk*)	ATP-regeneration	AY826832	Full length	–
Oxoglutarate/malate translocator	Carbon transporter	AY826859	5′ Partial (IL)	–
Ferredoxin-NADP(+) reductase (*petH*)	Electron transfer	AY826853	5′ Partial	–
Acyl carrier protein (*acp*)	Fatty acid synthesis	AY826829	Full length	Rd, Cr, Di, Gl
Geranylgeranyl reductase/ hydrogenase	Isoprenoid modification	AY826855	Full length	–
1-Deoxyxylulose-5-phosphate synthase (*dxs*)	Isoprenoid synthesis	AY826876	5′ Partial	–
Beta-keto-acyl reductase	Lipid synthesis	AY826869	Full length	–
Translation elongation factor Ts (*tsf*)	Plastid translation	AY826878	5′ Partial (IL)	Rd, Cr
Lipoate protein ligase	Protein modification	AY826879	5′ Partial (IL)	–
GAPDH	PS-dark	AY884246-7	Full length	–
Phosphoribulokinase (*prk*)	PS-dark	AY826860	Full length	–
Ribose-5-phosphate isomerase (*rpiA*)	PS-dark	AY826893	Full length	–
Fructose-1,6-bisphosphate aldolase (*fbaA*)	PS-dark	AAV71135	Full length	–
Phosphoglycerate kinase (*pgk*)	PS-dark	AY826862	5′ Partial (IL)	–
Transketolase (*tktA*)	PS-dark	AY826896	Full length	–
Carbonic anhydrase (*yadF*)	PS-dark	AY826838-40	Full length	–
RuBisCO form II	PS-dark	AY826897	Full length	–
Oxygen evolving enhancer 1 (*psbO*)	PS-light	AAM77465	Full length	–
Photosystem II extrinsic protein (*psbU*)	PS-light	AY826889	5′ Partial (IL)	–
Cytochrome f (*petA*)	PS-light	AY826881	5′ Partial (IL)	Rd, Cr, Di, Gl, Pl
ATP synthase subunit gamma (*atpC*)	PS-light	AY826835	Full length	–
Cytochrome b6 (*petC*)	PS-light	AY826843	5′ Partial	–
Cytochrome c6 (*petJ*)	PS-light	AY826872, AY884248	Full length	Rd
Photosystem I subunit XI (*psaL*)	PS-light	AY826885	Full length	Rd, Cr, Di
Ferredoxin (*petF*)	PS-light	AY826847-8	Full length	Rd, Cr, Di, Gl
Photosystem I protein E (*psaE*)	PS-light	AY826882	5′ Partial (IL)	Rd, Cr, Di, Gl
ATP synthase subunit C (*atpH*)	PS-light	AY826871, AY884249-55	Full length	Rd, Cr, Di, Gl, Eu, Pl
Cytochrome b559 (*psbF*)	PS-light	AY826887	Full length	Rd, Cr, Di, Gl, Eu, Pl
Photosystem II protein L (*psbL*)	PS-light	AY826888	Full length	Rd, Cr, Di, Gl, Eu, Pl
Light-harvesting protein	PS-light	AY826903	Full length	–
Light-harvesting protein	PS-light	AY826898-9	Full length	–
Light-harvesting protein	PS-light	AY826901	Full length	–
Light-harvesting protein	PS-light	AY826902	Full length	–
Light-harvesting protein	PS-light	AY826911	5′ Partial (IL)	–
Light-harvesting protein	PS-light	AY826900	5′ Partial (IL)	–
Light-harvesting protein	PS-light	AY826913	5′ Partial (IL)	–
Light-harvesting protein	PS-light	AY826912	5′ Partial (IL)	–
Light-harvesting protein	PS-light	AY826904	5′ Partial	–
Light-harvesting protein	PS-light	AY826905	5′ Partial	–
Light-harvesting protein	PS-light	AY826906	5′ Partial	–
Light-harvesting protein	PS-light	AY826907	5′ Partial	–
Light-harvesting protein	PS-light	AY826908	5′ Partial	–
Light-harvesting protein	PS-light	AY826909	5′ Partial	–
Light-harvesting protein	PS-light	AY826910	5′ Partial	–
Photosystem I subunit III (*psaF*)	PS-light	AY826884	5′ Partial (IL)	Rd, Cr, Di, Gl
Protochlorophyllide reductase subunit (*chlL*)	PS-light	AY826880	5′ Partial (IL)	Rd, Cr, Di, Gl, Eu
Ascorbate peroxidase	Reactive oxg. protection	AY826833	5′ Partial (IL)	–
Dimethyladenosine synthase (*ksgA*)	rRNA modification	AY826874	Full length	–
Queuine tRNA ribosyltransferase (*tgt*)	tRNA modification	AY826892	5′ Partial	–
Thylakoid 11 kDa protein	Unknown	AY826895	5′ Partial	–

52 proteins (sorted by pathway) are predicted from 63 cDNAs including isoforms. Sequences with incomplete leader sequences are indicated 5′ partial (IL). Rd, proteins that are plastid-encoded in red algae; Cr, cryptomonads; Di, diatoms; Gl, glaucophytes; Eu, euglenids; Pl, plants. Gene names are in brackets.

Table 2. Amino acid biases in *H. triquetra* presequence domains

	%Ala	%Ser	%Arg	%Thr	%Gln
SP-plastid	28.4	8.6	4	5.3	1.4
SP-secreted	24.2	7.4	5.3	2.6	1.6
TP N terminus – TM (class I)	25.5	8.1	4.5	7.4	11.2
TP N terminus – no TM (class II)	14.4	7.7	9.8	5.2	9.8
N terminus – mature secreted	10.4	9.6	3.8	6.5	2.7
TM	34.4	10.2	0.4	4.3	0
Post TM	24.1	4.5	16.7	5.8	3.9
Mature protein	9.6	5.1	4.2	4.9	3.4

SP, signal peptide; TP, transit peptide; TM, transmembrane domain.

analysis). The *H. triquetra* proteins were only deemed to be full length if (1) they had an N-terminal extension following the first methionine compared to the N terminus of cyanobacterial proteins, and (2) this presequence contained a predicted signal peptide (using the Hidden Markov model (HMM) method of SignalP 3.0[23]), consistent with targeting to secondary plastids in all other systems. For all 24 proteins, additional peptide sequence occurred between the predicted signal peptide cleavage site and the predicted mature protein, consistent with the presence of a transit peptide. Overall, therefore, each plastid-targeted protein was divided into three functional domains: SP, TP and mature protein.

We also assembled a set of ten proteins believed to be secreted, vacuolar, or remain in the ER in order to compare the targeting signals for plastid proteins with those of proteins that sort to different destinations within the endomembrane system (collectively designated "secreted" proteins: accession numbers AY826914–23). *H. triquetra* signal peptides were 15–45 amino acid residues (23 mean) in length. Comparison between plastid proteins and secreted proteins by amino acid composition showed no distinguishing features between these two groups (Table 2, Figure 1). They are both reduced in charged residues and strongly enriched in hydrophobic residues, particularly alanine (28.4% and 24.2%, respectively), compared to mature proteins.

Multiple transit peptide types in *H. triquetra*

The few dinoflagellate transit peptides previously described are reported to be similar to those of euglenids.[7] The amino terminus is enriched in the hydroxylated amino acids serine and threonine, as are plant chloroplast transit peptides, but downstream of this is a unique hydrophobic domain, shown in *Euglena* to function as a stop-transfer membrane anchor that is critical to the unique form of targeting they are proposed to share.[8–11] We searched for putative transmembrane (TM) helices downstream of the SP in the *H. triquetra* sequences using TMpred[24] (other prediction software was also used and gave the same results: not shown). In 15 of the 24 sequences, a TM helix was detected before the predicted start of the mature protein (Figure 2(a)). Immediately downstream of this helix is a strongly negatively charged domain, (particularly enriched for arginine)

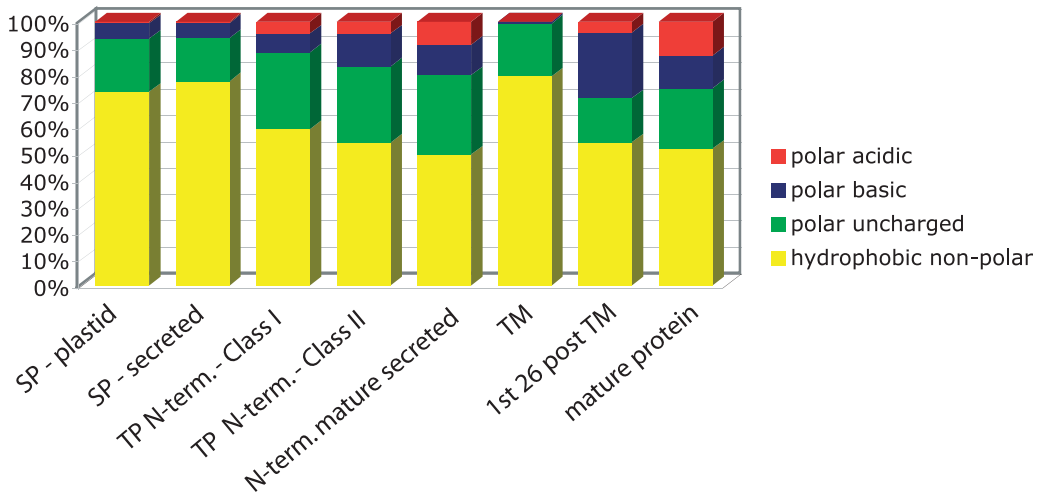

Figure 1. Amino acid composition of targeting peptides. Percentage bar charts of amino acid types in the peptide domains associated with plastid targeting in *H. triquetra*. Data for secreted proteins and the mature region of plastid proteins is also shown. Red, polar acidic residues (D, E); blue, polar basic residues (H, K, R); green, polar uncharged residues (C, N, Q, S, T, W, Y); yellow, hydrophobic non-polar residues (A, F, G, I, L, M, P, V). SP, signal peptide; TP, transit peptide; N-term., N terminus post SP cleavage (26 residues or less); TM, transmembrane domain.

consistent with the features of a stop-transfer membrane anchor[25,26] (Figure 2(a)). This is consistent with the expectation for all dinoflagellate plastid-targeted proteins, but seven of the 24 *H. triquetra* transit peptides did not contain these features (Figure 3(a)). It is unlikely that the predicted SP at the N terminus is in fact the TM domain of these seven proteins, because there is no negatively charged, arginine-rich region immediately downstream, as would be expected of the TM domain. Moreover, the hydropathy signal of post cleavage region of both sets of SPs is remarkably similar, and unlike that following the TM domain (Figures 2(a) and 3(a)).

The remaining two proteins of the 24 (cytochrome *c*6 and PsbO) are predicted to contain a TM helix (Figure 4(a)), but in a region that aligned with the cyanobacterial thylakoid signal peptide. Cyanobacteria use prokaryotic SPs to direct proteins to the thylakoids,[26] and this signal is thought to be conserved in targeting to the thylakoids of plastids. Both cytochrome *c*6 and PsbO are targeted into the thylakoid and are known to require this extra sorting element in other systems.[27,28]

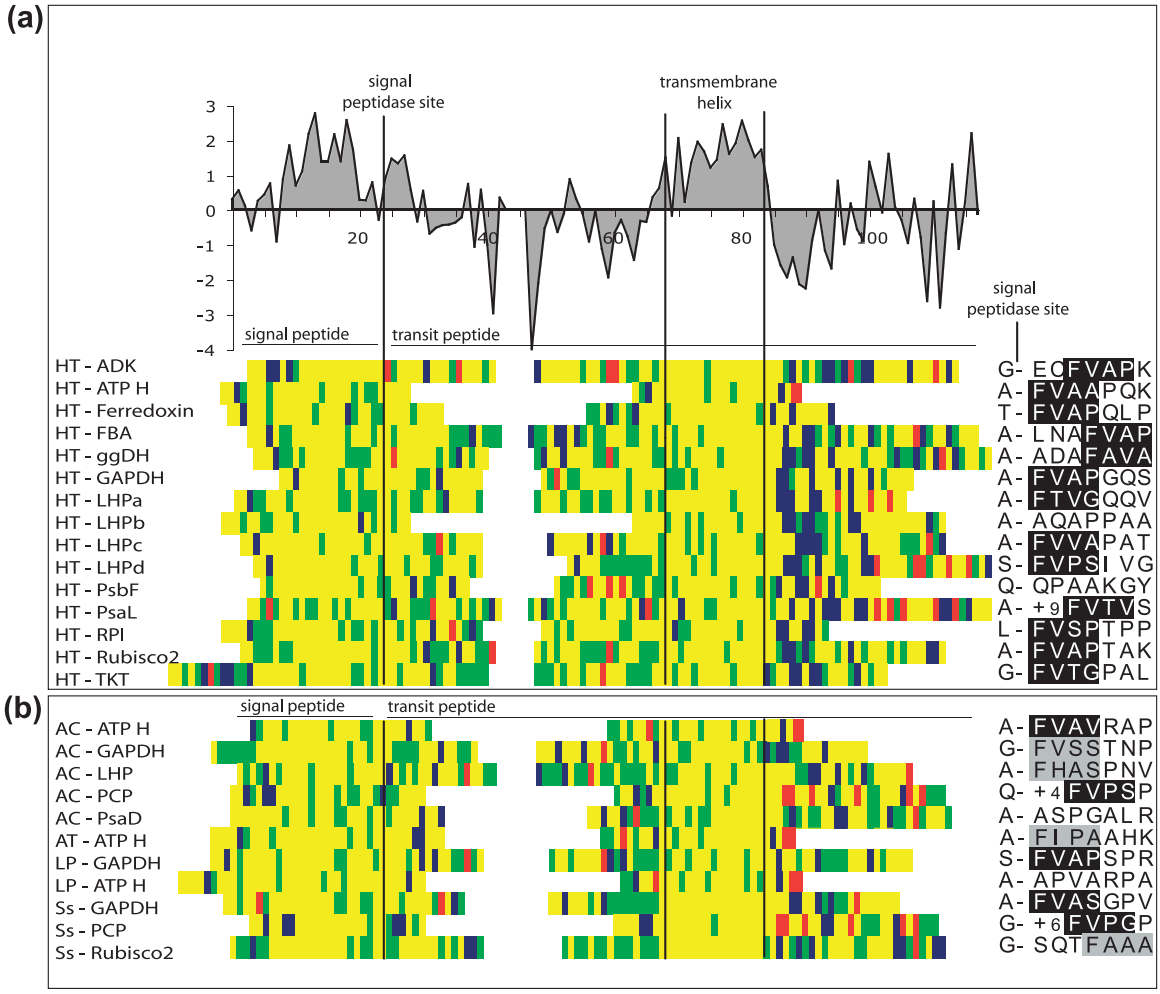

Figure 2. Class I transit peptides. Dinoflagellate plastid-targeting presequences that contain a hydrophobic transmembrane domain in the transit peptide (class I). (a) *H. triquetra* protein sequences (15) are represented by amino acid groups (red, polar acidic; blue, polar basic; green, polar uncharged; yellow, hydrophobic non-polar) and aligned at the predicted signal peptide cleavage site, and the predicted transmembrane domain. Average Kyte–Doolittle hydropathy scores for each column of the alignment are plotted above indicating the hydrophobic signal peptide and transmembrane domain, the relatively hydrophilic region in between, and the strongly hydrophilic (basic) region following the transmembrane domain. The position of a FVAP-motif is shown, relative to the signal peptidase site, shaded black where it consists of a F followed by three residues, one of which is valine, the remainder consisting of no more than one polar uncharged residue, and no acidic or basic residues. Motifs that fail these criteria by only one residue have a gray background. (b) Alignment of plastid proteins from other dinoflagellates that also have a transmembrane domain in their transit peptides. HT, *Heterocapsa triquetra*; AC, *Amphidinium carterae*; AT *Alexandrium tamarense*; LP, *Lingulodinium polyedrum*; Ss, *Symbiodinium* sp.; ADK, adenylate kinase; ATP H, ATP synthase subunit C; FBA, fructose 1,6 bisphosphate aldolase; ggDH, geranylgeranyl dehydratase/reductase; LHPa, AY826901; LHPb, AY826898; LHPc, AY826902; LHPd, AY826903; PsbF, cytochrome b559; PsaL, photosystem I subunit XI; RPI, ribulose phosphate isomerase; TKT, transketolase; PCP, Peridinin chlorophyll a protein; PsaD, photosystem I subunit II.

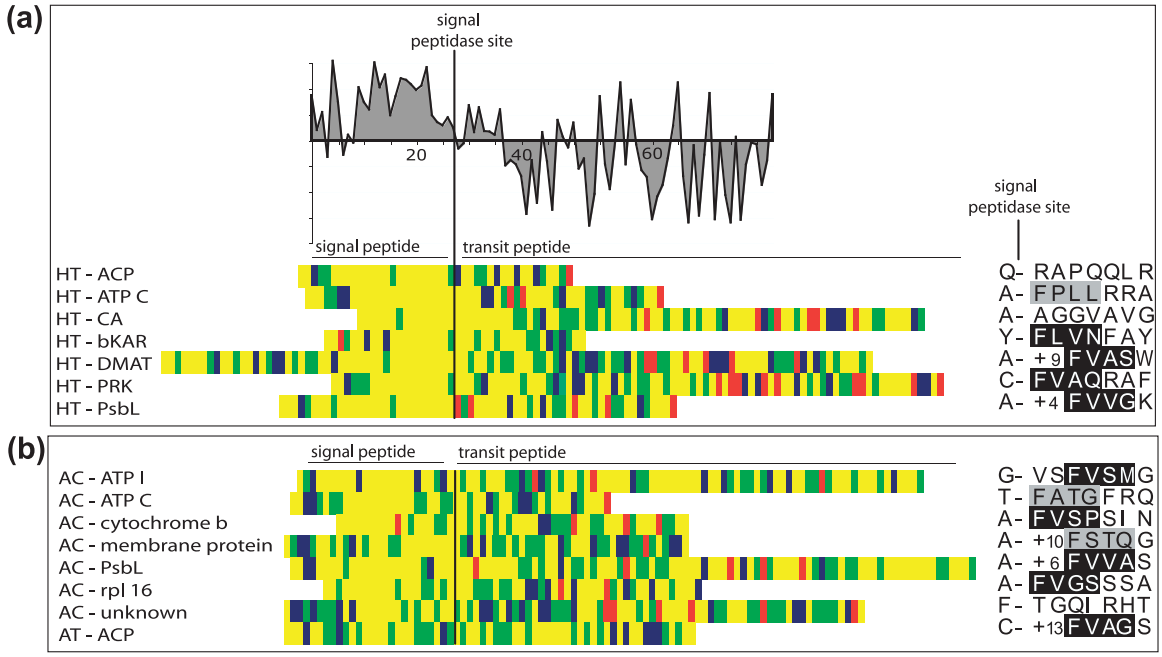

Figure 3. Class II transit peptides. Dinoflagellate plastid-targeting presequences lacking a hydrophobic transmembrane domain in the transit peptide (class II). *H. triquetra* (a) and other dinoflagellate (b) sequences are aligned as for Figure 2. ACP, acyl carrier protein; ATP C, ATP synthase subunit gamma; CA, carbonic anhydrase; bKAR, beta keto-acyl reductase; DMAT, dimethyladenoxine synthase; PRK, phosphoribulose kinase; PsbL, photosystem II protein L; ATP I, ATP synthase subunit I; rpl 16, ribosomal protein L16.

We therefore conclude that there are three types of TP in *H. triquetra* that are distinguished by either having a stop-transfer membrane anchor domain (class I), lacking an additional hydrophobic domain (class II), or possessing a thylakoid-targeting signal peptide (class III).

A common N terminus amongst transit peptide classes in *H. triquetra*

The amino acid composition was determined for the N-terminal 26 residues (or less if the mature protein or TM domain occurred sooner) of all class I and II transit peptides. Class III TP were excluded given the small sample size (two) and that the boundary between the TP and the thylakoid SP was not clear. Compositions were compared to the other presequence domains (SP and TM), as well as the mature proteins and the first 26 residues after the SP of the secreted proteins to see if any trend emerged. The N termini of both class I and II TPs were united by a number of common features (Figure 1, Table 2). Polar non-charged residues (green) contribute to the hydrophilicity of these domains, which also are reduced in acidic residues (4.5% and 3.6%, respectively) compared to the mature proteins (12.9%). Basic residues are not so reduced (7.2% and 13.7%), resulting in a net positive charge. Glutamine is relatively rare in mature proteins (3.4%), but is conspicuously abundant in TP regions (11.2% and 9.8%). Serine and alanine levels are elevated across the whole presequence, while threonine, previously noted to be high in dinoflagellates[7] and plant[5] TPs, was not especially elevated in *H. triquetra*. Hsp70 binding motifs are hypothesized to play a role in protein targeting to plastids in plants and the apicomplexan *Plasmodium falciparum*,[29,30] but were equally abundant in *H. triquetra* transit peptides

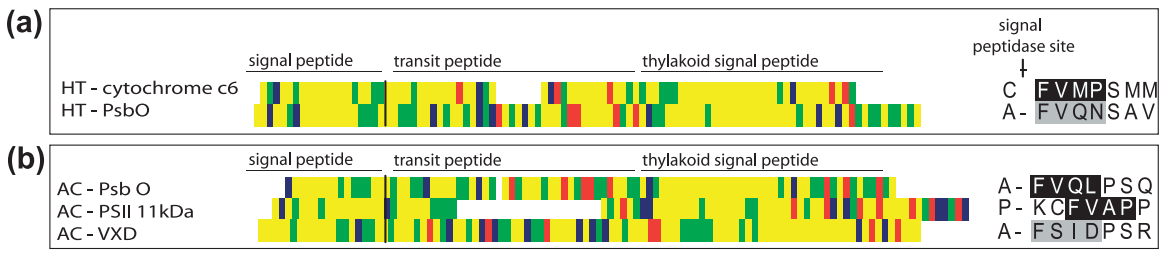

Figure 4. Class III transit peptides. Dinoflagellate plastid-targeting presequences that contain a thylakoid-targeting signal peptide in the transit peptide (class III). *H. triquetra* (a) and other dinoflagellate (b) sequences are aligned as for Figure 2. PsbO, oxygen evolving enhancer 1; PSII 11 kDa, photosystem II 11 kDa protein; VXD, violaxanthin de-epoxidase.

and mature proteins, and so are not a distinguishing feature of *H. triquetra* transit peptides.

Perhaps the most conspicuous feature of the 26-residue TP domain is a phenylalanine at or close to the N terminus in 18 of 22 proteins (Figures 2(a) and 3(a)). In *H. triquetra* the phenylalanine is typically followed by a relatively hydrophobic domain of three residues, one of which is valine (in all but one case), the remainder typically consisting of no more than one polar uncharged residue, and no acidic or basic residues (indicated with a black background in Figures 2 and 3 when all of these conditions occur). In five cases this sequence is FVAP. This motif does not seem to be linked to the presence of a signal peptide as only one of the ten secretory proteins has a phenylalanine at the first position after cleavage and only one has a "FVAP"-type motif within the first 26 residues of the cleavage point.

The shared features of the N terminus of class I and II TPs (the two class III proteins also conform to these features) suggest that this part of the TP is functionally conserved. Using these features, we defined criteria to discriminate between secreted proteins and those potentially targeted to the plastid. These criteria are for the first 26 residues post cleavage of a predicted signal peptide (or less if a TM domain is present sooner): (1) a phenylalanine occurs in the context of a FVAP-type motif; (2) this same region contains no more than 7.7% acidic residues (this allows for no more than two in a 26 residue stretch); and (3) a 10% or greater content of glutamine in this region. All of the predicted plastid proteins satisfy two or more of these criteria except PsbF, while nine of the ten secretory proteins are rejected using this cut-off of two criteria (Hsp90 satisfying two criteria). The criteria are not suggested to be definitive, but could identify putative plastid-targeted proteins for further study, so we screened the ESTs for open reading frames 60 residues or more that are predicted to possess an N-terminal signal peptide followed by sequence that meets two or more of the TP criteria. In addition to the predicted plastid-targeted proteins we recovered a further 28 sequences (Supplementary Table S1). Of these, 26 represent hypothetical proteins, 12 of which have homologs in other dinoflagellates. The other two cDNAs have significant matches to known proteins: one to bacterial phosphoenolpyruvate synthase and the other an RNA helicase. Interestingly, the RNA helicase is homologous to one predicted to be targeted to the apicoplast in *Plasmodium falciparum*.[31]

FVAP is an ancient targeting motif

Historically, plastid-targeting TP properties have been defined in plants, but recent large-scale sequencing has provided a broader representation of plastid proteins from diverse plastid types. The general features of a net positive charge and a bias for hydroxylated residues have held, albeit with some exceptions,[32] proving plants to be a good model overall. The phenylalanine motif that we observe in *H. triquetra*, however, is not a feature of plant TPs. Phenylalanine residues have also have been reported as conspicuous features of TPs in other chromalveolates, including apicomplexa,[32] the diatoms *Thalassiosira pseudonona* and *Phaeodactylum tricornutum*,[33,34] and cryptomonads.[35] Like *H. triquetra*, the *P. tricornutum* phenylalanine-motif is not associated with other endomembrane-targeted proteins and mutating the phenylalanine has been shown to disrupt plastid-targeting once proteins enter the endomembrane system.[33] TPs from organisms with secondary plastids derived from green algae (chlorarachniophyte *Bigelowiella natans* and euglenophyte *Euglena*) do not have elevated phenylalanine in the equivalent positions.[36] This dichotomy suggests that TPs have inherited characters according to the source of the plastid: red *versus* green.

We examined the source of this dichotomy: primary red and green plastids. From 47 red algal plastid-targeted proteins assembled from *Cyanidioschyzon merolae*[37] and other red algae, 91% contain a phenylalanine in the first ten residues, and 80% were in a hydrophobic context with none of the following three residues either acidic or basic (Supplementary Table S2). Similarly, from the 32 plastid-targeted proteins in the red algal endosymbiont nucleus of the cryptomonad *Guillardia theta*,[38] 88% contain a phenylalanine in the first ten residues, and 81% are followed by three uncharged residues (Supplementary Table S2). We compared this to the 1501 proteins currently annotated as being chloroplast-targeted by the Arabidopsis Information Resource† and only 26% had a phenylalanine in the first ten residues, and only 8.8% were in the hydrophobic context as described above. Interestingly, a similar motif has also now become apparent in plastid-targeted proteins from the third group with a primary plastid, glaucophytes.[39] The functional significance of this division is unclear, but it appears to be an ancient and highly conserved difference that has been maintained through subsequent symbioses, and even the remodeling of transit peptides and perhaps trafficking routes, such as in dinoflagellates. Moreover, the presence of the motif in glaucophytes suggests that it is ancestral to all three primary plastid-bearing lineages, but has been lost in the green lineage.

Transit peptide classes are conserved in dinoflagellate homologs

To investigate the conservation of the three classes of TPs amongst dinoflagellates other than *Heterocapsa*, we examined the leaders of the 22 available full-length dinoflagellate sequences (representing 15 proteins and four taxa),[7,17,18] and determined which class of TP each possessed

† http://arabidopsis.org/

(Figures 2(b), 3(b) and 4(b)) (note: for some thylakoid proteins, the presence of a thylakoid SP is unclear, hence these proteins are presently not designated class III). For all proteins for which there were representatives in more than one taxon, the class of TP was conserved. Class I proteins were represented in multiple taxa (up to four taxa per protein) by four proteins (GAPDH, ATP H, RuBisCO, and light harvesting proteins, Figure 2), class II by three proteins (ACP, ATP C and PsbL, Figure 3) and class III by one protein (PsbO, Figure 4). This apparently strict conservation of classes implies that the class of presequence may be difficult to alter, and may be significant for the targeting of any given plastid protein.

The features of the TP N terminus also show general conservation amongst these other dinoflagellates, with the exception of the elevated frequency of glutamine. Indeed, if the TP criterion 3 (above) is amended to a 10% requirement for serine in place of glutamine, all but two of the non-*Heterocapsa* proteins meet the criteria. This difference is unsurprising given that apicomplexan TPs from *Toxoplasma* and *Plasmodium* also show different amino acid preferences despite their TPs being interchangeable.[32,40]

Multiple targeting pathways for dinoflagellate plastids?

The presence of distinct classes of plastid-targeting presequences in dinoflagellates is novel in eukaryotes, and further complicates speculation on possible mechanisms of import into its three membrane-bound plastids. Nassoury *et al*. have proposed a targeting route based on their analysis of two proteins (RuBisCO and PCP) that parallels the route described for euglenids.[7] That is, proteins move from the ER to plastids *via* the Golgi apparatus while anchored in the membrane by the TM domain. This may be true for class I proteins (and all euglenid proteins studied to date), but what of the class II proteins lacking the critical TM domain? The other conserved features of the TP N terminus suggest that the same Toc/Tic apparatus is engaged, but do these proteins follow different trafficking routes to the plastid?

It is conceivable that elements of the mature protein could fulfill targeting functions in some plastid-targeted proteins. However, several of the dinoflagellate class II proteins do not contain TM domains in the mature proteins that could conceivably function as stop-transfer membrane anchors. It is therefore likely that these proteins are fully imported into the ER lumen during translation. Similarly, class III proteins that contain a thylakoid SP could use this motif to arrest protein import at the ER. It is also possible that with increased sampling we will find proteins that contain both a stop-transfer membrane anchor and a thylakoid SP, as has been reported from *Euglena*.[28] Nassoury *et al*. speculate that plastid proteins that span the membranes of their trafficking vesicles can be more easily retrieved if they are mistakenly delivered to the plasma membrane.[7] It is possible also that the cytosolic portion of plastid proteins may be involved in recruiting SNARE-type membrane targeting molecules, thus directly playing a part in their sorting to the plastid outer membrane. Neither of these explanations, however, account for class II proteins that most likely do not span the membranes of the sorting vesicles.

McFadden[1] suggests that plastid-targeting presequences for secondary plastids recapitulate the evolutionary history of the plastid. That is, transit peptides originated when genes relocated to the nucleus during the process of primary endosymbiosis, and SPs were added as a result of secondary endosymbiosis. Could the dichotomy between presequences in dinoflagellates also elude to some event in dinoflagellate evolution, such as the loss of a plastid membrane? If so, then genes already encoded in the nucleus of the red algal endosymbiont might encode class II transit peptides, while genes that moved more recently from the plastid genome to the dinoflagellate nucleus might have acquired novel features. However, genes that relocated to the nucleus exclusively in dinoflagellates (ATP H, cytochrome b559 and PsbL) include both class I and II categories. Conversely, genes encoded in the nucleus of all photosynthetic groups, presumably ancient transfers, also include both classes I and II TPs, suggesting no link between class I and II leaders and the evolutionary history of gene transfer.

Another explanation for the different presequence categories is that they might define some functional groups or final destinations within the plastids, since the conservation of presequence classes across dinoflagellate taxa implies some selection for the maintenance of class distinctions. However, no such distinctions are obvious from the current data: proteins of both class I and II function in the same pathways (photosynthesis, light and dark reactions), the same plastid locations (integral membrane proteins of the thylakoids such as LHPs, PsbL, PsaL; stromal proteins ACP, ferredoxin), and even the same complexes (ATPase subunits ATP H and ATP C). Altogether, the underlying reason for the distribution of class I and II leaders is not clear at present, so it may reflect something relating to trafficking rather than the ultimate function of the proteins. For example, if some proteins were inactivated by passage through the lumen of the Golgi while others were not, class I leaders may arise simply as a way to exclude sensitive proteins from this environment. Thus, the distinction between class I and II leaders could reflect physical characteristics of the proteins that use them rather than the role or location in the plastid of those proteins.

The different plastid-targeting presequence types observed in dinoflagellates add a new conundrum to the subject of protein sorting in plastid evolution. Dinoflagellate plastids appear to have originated with those of other chromalveolates, so there has

been substantial modification of plastid protein trafficking routes in the dinoflagellate lineage: a membrane was lost and the nature of at least some of the presequences changed. If the membrane loss was the driving force for these changes, as others have speculated,[7] then it will be interesting to see if euglenids have equivalent class divisions in their plastid presequences (to date no homologs of the dinoflagellate class II proteins have been characterized in euglenids). The establishment of a targeting system is considered one of the major challenges in the foundation of a new plastid, so the apparently duality of dinoflagellate targeting also raises interesting questions about how this system can be changed after the plastid is integrated.

Methods

Library construction

H. triquetra (CCMP 449) was cultivated in Guillard's f2-Si medium[41] at 16 °C, with a 12 hour/12 hour light/dark cycle. Cells were harvested in batches throughout the light cycle and RNA prepared using TRIzol® Reagent (Invitrogen) according to the manufacturer's instructions. Pooled RNA was used to construct a cDNA library in the vector PCDNA3.1 (+) (DNA Technologies, Inc., USA) that was transformed into *Escherichia coli*.

EST sequencing

A total of 9309 5′-end sequenced ESTs were analyzed by PEPdb† for automatic quality and vector-trimming, and assembled into contiguous sequence clusters. Proteins were identified based on sequence similarity searches against public databases. In addition, proteins representing known plastid functions from other eukaryotes and cyanobacteria, were used to search *H. triquetra* data, resulting in a set of putative plastid-targeted protein genes which were completely sequenced on both strands from over-lapping cDNA clones for each cluster. Protein alignments to homologs from other organisms were made using Clustal X.

Data Bank accession codes

Sequence data from this article have been deposited with the GenBank data library under accession number AY826826-947, AY884246-55.

Acknowledgements

This work was supported by the Protist EST Program of Genome Canada/Genome Atlantic, and by a grant (MOP-42517) from the Canadian Institutes for Health Research (CIHR). We thank B. Green for comments on the manuscript. R.F.W. is supported by Fellowships from CIHR and the Michael Smith Foundation for Health Research (MSFHR), J.M.A. was supported by postdoctoral fellowships from CIHR and the Killam Foundation and P.J.K. is a scholar of the CIAR and a new investigator of the CIHR and MSFHR.

† http://amoebidia.bcm.umontreal.ca/pepdb/searches/welcome.php

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.03.030

References

1. McFadden, G. I. (1999). Plastids and protein targeting. *J. Eukaryot. Microbiol.* **46**, 339–346.
2. Keeling, P. (2004). A brief history of plastids and their hosts. *Protist*, **155**, 3–7.
3. Martin, W. & Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118**, 9–17.
4. van Dooren, G. G., Schwartzbach, S. D., Osafune, T. & McFadden, G. I. (2001). Translocation of proteins across the multiple membranes of complex plastids. *Biochim. Biophys. Acta*, **1541**, 34–53.
5. Bruce, B. D. (2001). The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim. Biophys. Acta*, **1541**, 2–21.
6. Jarvis, P. & Soll, J. (2002). Toc, tic, and chloroplast protein import. *Biochim. Biophys. Acta*, **1590**, 177–189.
7. Nassoury, N., Cappadocia, M. & Morse, D. (2003). Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J. Cell Sci.* **116**, 2867–2874.
8. Sulli, C., Fang, Z., Muchhal, U. & Schwartzbach, S. D. (1999). Topology of Euglena chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J. Biol. Chem.* **274**, 457–463.
9. Sulli, C. & Schwartzbach, S. D. (1996). A soluble protein is imported into Euglena chloroplasts as a membrane-bound precursor. *Plant Cell*, **8**, 43–53.
10. Sulli, C. & Schwartzbach, S. D. (1995). The polyprotein precursor to the Euglena light-harvesting chlorophyll a/b-binding protein is transported to the Golgi apparatus prior to chloroplast import and polyprotein processing. *J. Biol. Chem.* **270**, 13084–13090.
11. Osafune, T., Schiff, J. A. & Hase, E. (1991). Stage-dependent localization of LHCP II apoprotein in the Golgi of synchronized cells of Euglena gracilis by immunogold electron microscopy. *Expt. Cell Res.* **193**, 320–330.
12. Leander, B. S. (2004). Did trypanosomatid parasites have photosynthetic ancestors? *Trends Microbiol.* **12**, 251–258.
13. Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol. Rev. Camb. Philos. Soc.* **73**, 203–266.
14. Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* **18**, 418–426.
15. Patron, N. J., Rogers, M. B. & Keeling, P. J. (2004). Gene replacement of fructose-1,6-bisphosphate

aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot. Cell*, **3**, 1169–1175.
16. Harper, J. T. & Keeling, P. J. (2003). Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol. Biol. Evol.* **20**, 1730–1735.
17. Bachvaroff, T. R., Concepcion, G. T., Rogers, C. R., Herman, E. M. & Delwiche, C. F. (2004). Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist*, **155**, 65–78.
18. Hackett, J. D., Yoon, H. S., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E. *et al*. (2004). Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr. Biol.* **14**, 213–218.
19. Zhang, Z., Green, B. R. & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature*, **400**, 155–159.
20. Howe, C. J., Barbrook, A. C., Koumandou, V. L., Nisbet, R. E., Symington, H. A. & Wightman, T. F. (2003). Evolution of the chloroplast genome. *Philos. Trans. R. Soc. London B Biol. Sci.* **358**, 99–106 [discussion 106–7].
21. Ishida, K. & Green, B. R. (2002). Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc. Natl Acad. Sci. USA*, **99**, 9294–9299.
22. Tanikawa, N., Akimoto, H., Ogoh, K., Chun, W. & Ohmiya, Y. (2004). Expressed sequence tag analysis of the dinoflagellate Lingulodinium polyedrum during dark phase. *Photochem. Photobiol.* **80**, 31–35.
23. Bendtsen, J. D., Nielsen, H., Heijne, G. v. & Brunak, S. (2004). Improved prediction of singal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795.
24. Hofmann, K. & Stoffel, W. (1993). TMbase: A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, **374**, 166.
25. Kuroiwa, T., Sakaguchi, M., Mihara, K. & Omura, T. (1991). Systematic analysis of stop-transfer sequence for microsomal membrane. *J. Biol. Chem.* **266**, 9251–9255.
26. Mackle, M. M. & Zilinskas, B. A. (1994). Role of signal peptides in targeting of proteins in cyanobacteria. *J. Bacteriol.* **176**, 1857–1864.
27. Chaal, B. K., Ishida, K. & Green, B. R. (2003). A thylakoidal processing peptidase from the heterokont alga Heterosigma akashiwo. *Plant Mol. Biol.* **52**, 463–472.
28. Vacula, R., Steiner, J. M., Krajcovic, J., Ebringer, L. & Loffelhardt, W. (1999). Nucleus-encoded precursors to thylakoid lumen proteins of Euglena gracilis possess tripartite presequences. *DNA Res.* **6**, 45–49.
29. Ivey, R. A., 3rd, Subramanian, C. & Bruce, B.D (2000). Identification of a Hsp70 recognition domain within the rubisco small subunit transit peptide. *Plant Physiol.* **122**, 1289–1299.
30. Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S. *et al*. (2003). Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. *Science*, **299**, 705–708.
31. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W. *et al*. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, **419**, 498–511.
32. Ralph, S. A., Foth, B. J., Hall, N. & McFadden, G. I. (2004). Evolutionary pressures on apicoplast transit peptides. *Mol. Biol. Evol.* **21**, 2183–2194.
33. Kilian, O. & Kroth, P. G. (2005). Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J.* **41**, 175–183.
34. Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H. *et al*. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science*, **306**, 79–86.
35. Deane, J. A., Fraunholz, M., Su, V., Maier, U. G., Martin, W., Durnford, D. G. & McFadden, G. I. (2000). Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. *Protist*, **151**, 239–252.
36. Rogers, M. B., Archibald, J. M., Field, M. A., Li, C., Striepen, B. & Keeling, P. J. (2004). Plastid-targeting peptides from the chlorarachniophyte Bigelowiella natans. *J. Eukaryot. Microbiol.* **51**, 529–535.
37. Matsuzaki, M., Misumi, O., Shin, I. T., Maruyama, S., Takahara, M., Miyagishima, S. Y. *et al*. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature*, **428**, 653–657.
38. Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T. *et al*. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, **410**, 1091–1096.
39. Steiner, J. M. & Loffelhardt, W. (2002). Protein import into cyanelles. *Trends Plant Sci.* **7**, 72–77.
40. Waller, R. F., Reed, M. B., Cowman, A. F. & McFadden, G. I. (2000). Protein trafficking to the plastid of Plasmodium falciparum is *via* the secretory pathway. *EMBO J.* **19**, 1794–1802.
41. Andersen, R. A., Morton, S. L. & Sexton, J. P. (1997). CCMP-Provasoli-Guilard National Center for Culture of Marine Phytoplankton. *J. Phycol.* **33**, 1–75 [Supplement].

Edited by J. Karn

(Received 26 January 2005; received in revised form 9 March 2005; accepted 11 March 2005)