ELSEVIER

Biochimica et Biophysica Acta 1541 (2001) 34–53

BIOCHIMICA ET BIOPHYSICA ACTA

BBA

www.bba-direct.com

Review

Translocation of proteins across the multiple membranes of complex plastids

Giel G. van Dooren [a], Steven D. Schwartzbach [b], Tetsuaki Osafune [c], Geoffrey I. McFadden [a,*]

[a] *Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Melbourne, Vic. 3010, Australia*
[b] *Department of Microbiology and Molecular Cell Sciences, University of Memphis, Memphis, TN 38152, USA*
[c] *Department of Life Sciences, Nippon Sport Science University, Yokohama 227, Japan*

Received 13 August 2001; accepted 13 August 2001

Abstract

Secondary endosymbiosis describes the origin of plastids in several major algal groups such as dinoflagellates, euglenoids, heterokonts, haptophytes, cryptomonads, chlorarachniophytes and parasites such as apicomplexa. An integral part of secondary endosymbiosis has been the transfer of genes for plastid proteins from the endosymbiont to the host nucleus. Targeting of the encoded proteins back to the plastid from their new site of synthesis in the host involves targeting across the multiple membranes surrounding these complex plastids. Although this process shows many overall similarities in the different algal groups, it is emerging that differences exist in the mechanisms adopted. © 2001 Published by Elsevier Science B.V.

Keywords: Malaria; Protein transport; Endoplasmic reticulum; Vesicle; Golgi apparatus

1. Introduction

Imagine you are in a boat, on *the* ocean, about 600 million years ago. In the water around you an entire biosystem flourishes, with red and green algae at the base of a food web that provides nutrients for invertebrates like trilobites and top-of-the-foodchain predators like *Anomalocarus*. The importance of algae in these biosystems is due to their ability to obtain energy from the sun and convert this to sugar. The ability of these eukaryotic (nucleated) organisms to photosynthesize was acquired hundreds of millions years earlier [1], through a process called primary endosymbiosis.

Primary endosymbiosis arose when a non-photosynthetic eukaryote acquired the services of a photosynthetic cyanobacterium (Fig. 1). Current evidence, based largely on phylogenetic trees of algal proteins and preservation of operons, suggests that primary endosymbiosis was a singular event [2–4], with this primordial photosynthetic eukaryote diverging into the red algal and green algal/plant lineages. If, when standing in your boat 600 or so million years ago (the actual dates are not clear), you had a plankton filter and a decent enough microscope, you might witness the beginnings of a second wave of algal evolution. This wave, known as secondary endosym-

* Corresponding author.
E-mail address: g.mcfadden@botany.unimelb.edu.au (G.I. McFadden).

0167-4889/01/$ – see front matter © 2001 Published by Elsevier Science B.V.
PII: S0167-4889(01)00154-9

biosis, would produce much of the algal diversity witnessed today, ranging from organisms that cause toxic red tides and malaria, to those whose shells form the abrasive material that cleans your teeth every morning.

2. The evidence for secondary endosymbiosis

Our understanding of secondary endosymbiosis began in the late 1970s, with the publication of three seminal papers written on the topic by Sally Gibbs [5,6] and Jean Whatley and colleagues [7]; see also [8]. The argument that red algal and green algal/plant plastids were derived from cyanobacteria had largely been resolved with the finding that these plastids contained DNA that was similar to the DNA of cyanobacteria (e.g., [9]), although this plastid genome was greatly reduced. One other argument for the occurrence of primary endosymbiosis was that plastids were surrounded by two membranes, which resembled the two membranes surrounding cyanobacteria [10]. Gibbs [5] and Whatley et al. [7], however, noted that the plastids of *Euglena*, although morphologically similar to those of green algae, differed in one crucial respect: they were surrounded by three membranes rather than two. They interpreted this extra membrane as evidence for a secondary endosymbiosis, where a photosynthetic eukaryote (green alga) was engulfed by a eukaryotic host (Fig. 1). In

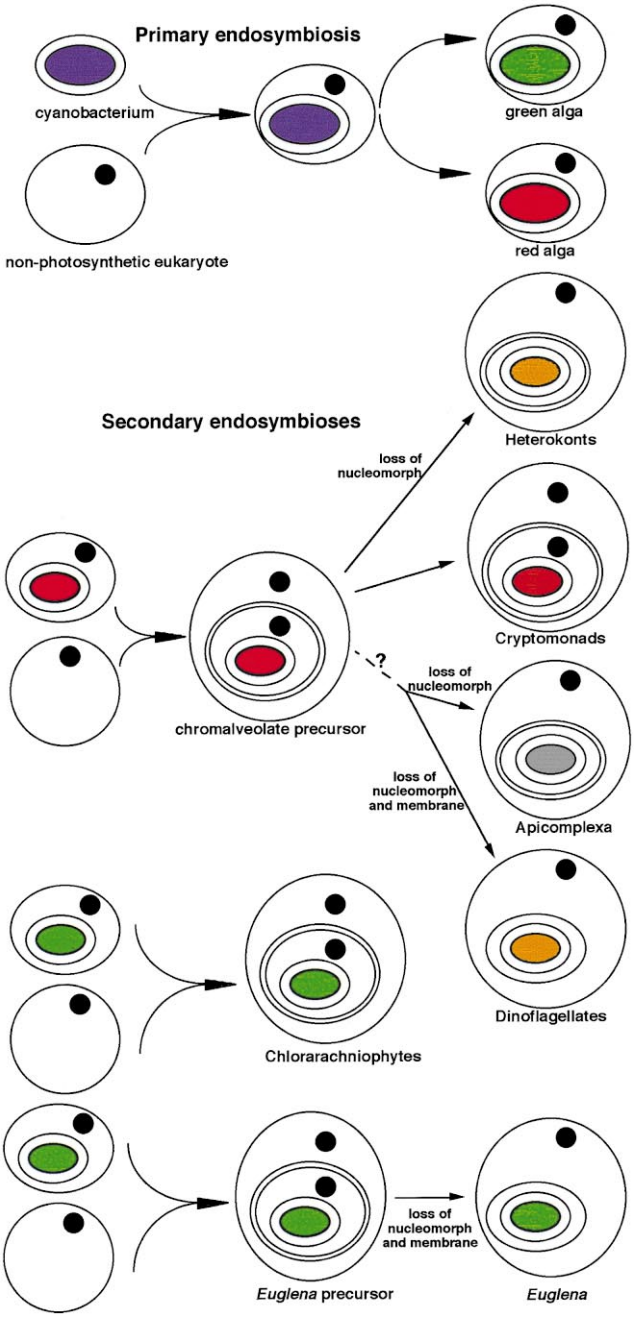


Fig. 1. Schema showing the evolution of plastids derived from primary and secondary endosymbioses. Primary endosymbiosis involved the incorporation of a cyanobacterium into a non-photosynthetic eukaryote. After this single primary endosymbiotic event, the phototrophic eukaryote diverged to produce the red and green algal lines, as well as the glaucocystophytes (not shown). Secondary endosymbiosis involved the incorporation of a phototrophic eukaryote into a non-photosynthetic eukaryote and evolved at least three separate times (see text for details). Soon after the establishment of the endosymbiosis, the host cell contained a plastid surrounded by four membranes, and a remnant nucleus (nucleomorph) located between the outer two and inner two membranes of the plastid. This nucleomorph is still found in cryptomonads and chlorarachniophytes. Consequent evolution involved the loss of the nucleomorph (as seen in all other organisms with secondary plastids) and the loss of one of the membranes surrounding the plastid (as seen in euglenoids and peridinin-containing plastids of dinoflagellates). This diagram indicates the three putative secondary endosymbioses that occurred, one leading to euglenoids, another to chlorarachniophytes, and a final one to the chromalveolates, which includes heterokonts, cryptomonads, apicomplexa and dinoflagellates [44]. The algal symbiont involved in the establishment of these endosymbioses (green or red) is shown. Alternative hypotheses posit that multiple separate endosymbioses produced dinoflagellates, cryptomonads, apicomplexa and heterokonts [47].

this schema, the third membrane represents either the plasma membrane of the green-algal symbiont [5], or the host cell vacuole membrane in which the symbiont lies [7]. The occurrence of three membranes surrounding the plastids was also observed in dinoflagellates, while the plastids of heterokont algae (diatoms, brown algae etc.), cryptomonads and haptophytes were all found to be surrounded by four membranes, again implying a secondary endosymbiosis had taken place (Fig. 1 and [6,7]).

To base an entire theory around the number of membranes that surround a plastid is tenuous at best. However, since the late 1970s morphological, phylogenetic and biochemical studies have strongly supported the theory of secondary endosymbiosis. Morphological observations of cryptomonad algae noted that the region between the outer two and inner two plastid membranes (called the periplastidal space) contained eukaryotic-like ribosomes [11,12]. Furthermore, the cryptomonad periplastidal space was found to contain a membrane-bound granular structure that resembles a nucleus, and was thus dubbed the nucleomorph [7,13]. It was shown that this nucleomorph indeed contains DNA [14,15], although much like the plastid genome of primary plastids, its gene content is greatly reduced [16,17]. Phylogenetic analysis of the two 18S rRNA subunits from cryptomonads [18] and subsequent localization of one of the genes to the nucleomorph [19,20] confirmed that the nucleomorph DNA was very different from that of the host cell, with the nucleomorph rRNA related to red algae. These findings suggested that the cryptomonad periplastidal space represents the cytosol of a red-algal symbiont, with the nucleomorph its vestigial nucleus. A nucleomorph has also been found in the periplastidal space of chlorarachniophyte algae [21], with phylogenetic evidence suggesting it is the vestigial nucleus of a green algal symbiont [22].

3. Secondary endosymbiosis: how and how often?

A popular explanation for the establishment of a secondary endosymbiosis is the 'stuck-in-the-throat' theory, which suggests that a phagotrophic, non-photosynthetic eukaryote swallows but fails to digest a photosynthetic eukaryote [23]. Over time, the association between these two organisms becomes permanent. Gradually, genes from the symbiont nucleus (nucleomorph) are either transferred to the host nucleus or lost, which may occur when the gene products are no longer required or when host nuclear-encoded proteins take over their function. The advantages of the symbiosis that leads to the development of an obligate relationship for the two organisms are not immediately clear. Although the word 'symbiosis' suggests mutual benefit, it is unclear whether the secondary endosymbiont was a captive, a commensal or an invader. Theoretically, the symbiont gains the advantage of being hidden from other predators. However, life as an endosymbiont restricts the opportunity for genetic recombination with other cells of its kind, and poses spatial problems when it comes to division. Conversely, the advantages for the host would seem obvious with the newfound ability to manufacture their own energy. However, it is unlikely that during the initial symbiosis, the symbiont freely gave up ATP or carbohydrates to the host cell. Indeed, having a whopping great alga stuck in your primordial throat is not conducive to further feeding for the host cell. In recent years, sophisticated metabolic models have been proposed for the evolution of eukaryotic cells and mitochondria [24,25], and it is possible that the evolution of secondary plastids also has a metabolic explanation. For example, Lee and Kugrens [26,27] propose that low environmental CO_2 levels 300 million years ago meant that algae located in the acidic food vacuoles of 'hosts' had more soluble CO_2 available to them.

Although the mechanisms remain unclear, the evolution of secondary endosymbioses raises two interesting, and to an extent inter-related, questions. How often did secondary endosymbiosis occur? And, given that the majority of plastid genes are located in the nucleus and translated in the cytosol, how are the proteins targeted across the multiple membranes to the plastid? Based on the number of membranes surrounding their plastids, seven phyla can be considered as having secondarily derived plastids (Table 1). Euglenoid and most dinoflagellate plastids are surrounded by three membranes (reviewed in [5]), while the plastids of heterokonts, haptophytes, cryptomonads and chlorarachniophytes are surrounded by four membranes (reviewed in [6,28]). A final group containing secondarily-derived plastids are the Api-

Table 1
Features of secondarily-derived plastids

Phylum	No. of plastid membranes	CER?	Nucleomorph?	Derivation
Heterokonts	4	Yes	No	Red alga
Haptophytes	4	Yes	No	Red alga
Cryptomonads	4	Yes	Yes	Red alga
Apicomplexa	4 (?)	No	No	Red alga
Dinoflagellates	3	No	No	Red alga
Chlorarachniophytes	4	No	Yes	Green alga
Euglenoids	3	No	No	Green alga

The first column shows the number of membranes surrounding plastids in each of the phyla thought to contain a secondarily-derived plastid. The second and third columns indicate the presence or absence of a chloroplast endoplasmic reticulum (CER) and nucleomorph, respectively. The final column indicates the type of alga from which the secondary plastid is most likely to have derived. See text for discussion.

complexa. Apicomplexa are a group of intra-cellular parasites that include *Plasmodium* spp., the causative agents of malaria, and the opportunistic human pathogen *Toxoplasma gondii*. This group has recently been shown to contains a non-photosynthetic plastid [29–31], dubbed the apicoplast, which probably functions in fatty acid and isoprenoid biosynthesis [32–35]. It appears that the plastids of this group are also surrounded by four membranes [36,37], although one study concluded that the plastid of *P. falciparum* is bounded by three membranes [38].

Evidence concerning the frequency of secondary endosymbiosis, and the phototrophic eukaryotes involved in the symbioses, comes from morphological and biochemical, as well as phylogenetic sources. The outer plastid membrane of cryptomonads, haptophytes and heterokonts bears ribosomes, and in most cases is continuous with the nuclear envelope through the rough endoplasmic reticulum (ER; reviewed in [39]). For these reasons it is known as the chloroplast endoplasmic reticulum (CER). All algae containing a CER have chlorophyll *c2* suggesting they either evolved from the same endosymbiotic event or harbour closely related symbionts [39]. Phylogenetic analysis of cryptomonad nucleomorph DNA unequivocally grouped cryptomonad endosymbionts with red algae [18]. Other analyses have grouped both diatoms and dinoflagellate endosymbionts with red algae [3,40], whereas chlorarachniophyte endosymbionts group with green algae [22]. The origin of the euglenoid plastid is less clear, with most recent studies placing them with the green algae (e.g., [3]; see review in [41]). Initially, apicomplexan plastids were grouped with green algae [31], but the gene order on the plastid genome [42,43] and their glyceraldehyde 3-phosphate dehydrogenase (GAPDH) phylogeny [44] suggests a red algal origin. With euglenoid and chlorarachniophyte plastids probably derived from green algae in separate events, and the other secondary plastids probably derived from red algae, there must have been at least three secondary endosymbiotic events. Indeed, some studies place host cells on entirely separate evolutionary branches, and this has been interpreted as evidence that secondary endosymbioses may have occurred on six or so separate occasions (reviews in [45–47]).

The question of how many secondary endosymbioses occurred is vexed by the fact that phylogenetic analyses are often hampered by the extreme divergence of rapidly-evolving plastid-encoded genes. For example, Zhang et al. [48] showed that two plastid genes of dinoflagellates formed a clade with the equivalent genes in apicomplexa. However, they tempered their conclusions by noting that the apparent grouping may be due to long-branch artefacts, caused by the rapid evolution in these plastid genomes. In a recent study, Naomi Fast and colleagues [44] partially overcame this obstacle by analysing the phylogenetic relationships between nuclear-encoded GAPDH genes from a range of organisms. Interestingly, they found that dinoflagellates, apicomplexa, heterokonts and cryptomonads all contained a plastid-targeted GAPDH gene that resembled not the plastid GAPDH genes of red and green algae, but instead the cytosolic GAPDH found in eukaryotes (for review see [49]). Furthermore, they found that these plastid-targeted GAPDH genes formed a well-supported clade, suggesting that after a duplication

event in cytosolic GAPDH, one copy became targeted to the plastid of these organisms. The most logical conclusion from this finding is that these secondarily-derived plastids share a common, so-called chromalveolate, origin, with the GAPDH duplication event having occurred before divergence of the different groups.

So what light does the preceding discussion of the evolution of secondary endosymbioses shed on the events that took place 600 or so million years ago. Although more evidence is required to support the common origin theory of Fast et al. [44], it suggests that at most three secondary endosymbiotic events occurred. One event led to the evolution of euglenoids, another to the chlorarachniophytes, and a final one to the chromalveolates (heterokonts, cryptomonads, dinoflagellates and apicomplexa; Fig. 1). Haptophytes also probably arose by secondary endosymbiosis of a red alga (Table 1), but their GAPDH genes have not been characterized [44] and their plastid DNA has received only limited studied. Cavalier-Smith [41] has argued that the plastids of euglenoids and chlorarachniophytes arose from the same symbiotic event, although evidence for this is weak. Due to the lack of mineralized cell walls in most of the above groups, dating the exact appearance of organisms containing secondary plastids is difficult. However, by ~400 million years ago (Mya) possible fossils of dinoflagellate cysts are observed (review in [50]), which puts the secondary endosymbiosis leading to those groups before this. By 185 Mya fossils of diatoms and haptophytes had all appeared (review in [27]), with the actual evolution of these groups also preceding their appearance in the fossil record. Cavalier-Smith [51] argues that secondary endosymbioses were established soon after primary endosymbiosis, around 600 Mya. This suggests that the acquisition of secondary plastids in the heterokont/dinoflagellate group facilitated a second major eukaryotic explosion, leading to the evolution of several of the world's major algal phyla. However, due to limited data and incomplete fossil evidence, this conclusion remains subject to revision.

Another conclusion to draw from the evolution of secondary endosymbioses is that the number of membranes surrounding the acquired plastid could be somewhat of a red herring when it comes to determining evolutionary relationships between the phyla. If Fast et al.'s interpretations [44] are correct, plastids surrounded by a CER are all related, as initially argued by Cavalier-Smith [52]. However, also related to this group are the Apicomplexa, whose plastids are also surrounded by four membranes, but lack a CER, and dinoflagellates, whose plastids are surrounding by three membrane and lack a CER (this suggests that aspects of the protein targeting mechanisms to the plastids of the dinoflagellate/heterokont group must have evolved after several of these lineages had diverged). Conversely, euglenoids are apparently unrelated to dinoflagellates, even though their plastids are surrounded by three membranes, and chlorarachniophytes are unrelated to apicomplexa, even though their plastids are surrounded by four membranes and appear to lack a CER.

4. Protein import into simple plastids with two envelope membranes

Our focus now turns to how nuclear-encoded proteins enter secondarily-derived plastids, which serves as the subject for the remainder of this review. To better understand how this process may occur, it is first necessary to review how nuclear-encoded proteins are targeted to primary plastids in red algae and green algae/plants, which are bound by two membranes. Although the major clue to the cyanobacterial origin of red and green algal plastids was the presence of a cyanobacterial-like genome in the organelle, it was apparent that this genome was greatly reduced. As in secondary endosymbiosis, most of the genes required for primary plastid function have been transferred to the cell nucleus. This means that most plastid proteins are translated in the cell cytosol, from where they must travel to their destination.

The transformation of a cyanobacterial endosymbiont into the primary (or 'simple') chloroplast of green algae and higher plants, thus, necessitated the evolution of a mechanism to transport nuclear-encoded, cytoplasmically-synthesized proteins across the two plastid envelope membranes. The general protein import pathway into simple chloroplasts is post-translational and dependent upon a transit peptide, a positively charged N-terminal region rich in hydroxylated amino acids, that nevertheless appears

to lack any primary sequence consensus (for detailed reviews of import into simple chloroplasts see [53,54] and articles in this volume). The outer and inner chloroplast envelopes contain protein complexes known as Toc (translocon of the outer chloroplast membrane) and Tic (translocon of the inner chloroplast membrane) that mediate translocation into the plastid [55]. In an irreversible ATP- and GTP-dependent process, the transit peptide interacts with Toc components and inserts through the outer envelope membrane translocation channel. After contacting the Tic machinery, import proceeds by ATP-dependent insertion of the precursor through the inner membrane translocation channel, followed by translocation into the stroma at envelope contact sites, regions of association between the Toc and Tic translocation machinery [53,54]. Finally a stromal peptidase removes the transit peptide [56]. Thylakoid lumen proteins have bipartite transit peptides, with a proximal stroma-targeting domain and a distal signal peptide-like domain. The latter functions within the chloroplast as a thylakoid targeting signal that is removed in the thylakoid lumen.

Transit peptides are found on proteins imported into the chloroplasts of green algae, red algae, higher plants and the cyanelles of glaucocystophytes. Nuclear-encoded cyanelle proteins [57] and red algal chloroplast proteins [58] are imported into isolated pea chloroplasts, suggesting that all organisms arising from the single primary endosymbiotic event use a common, transit peptide-mediated import pathway. Many components of the outer and inner membrane translocons have been cloned, and some have cyanobacterial homologues presumed to have ancestral functions in protein secretion [59–62]. Taken together, these results suggest that the transit peptide import system evolved through modification of existing membrane translocation systems present in the primary endosymbiont, with contributions from the host that induced the transport to operate in the reverse direction [59–62].

The additional membranes enclosing complex plastids complicate the protein import process. Gibbs realized this [6,39] and for plastids surrounded by a CER, suggested that chloroplast precursor proteins are translated on ribosomes attached to the CER, and co-translationally imported across this outermost membrane. Although based solely on morphological data, Gibbs' surmise turned out to be correct. Indeed we now believe that N-terminal signal peptides mediate protein transport via the secretory system across the outermost membrane of *all* secondary chloroplasts. This occurs either directly, as in CER-bearing plastids, or via the endomembrane system in systems where plastids have a smooth outer membrane. However, as this review examines protein import into the different types of complex chloroplasts, it will become evident that much remains to be learned about the mechanisms of protein translocation across the various inner membranes of these chloroplasts.

5. Protein import into plastids containing a CER

The first clues that ribosomes on the outer CER membrane were indeed involved in plastid protein translocation came from the sequences of nuclear-encoded, plastid-targeted proteins. cDNA sequences encoding several plastid-targeted proteins in heterokonts indicated the presence of a hydrophobic signal peptide-like region at the N-terminus [63–69]. Signal peptide-like regions also occur at the N-terminus of cryptomonad proteins [70,71]. Analysis of the region following the signal peptides of diatom and cryptomonad plastid-targeted proteins reveals a positively-charged region rich in hydroxylated residues that has the characteristic properties of a stromal targeting transit peptide (e.g., [70–72]). Bhaya and Grossman [73] demonstrated that the signal peptide region of a diatom plastid-targeted protein was capable of directing co-translational import of the protein into microsomal membranes in vitro, with the leader sequence being cleaved upon import. More recently, Wastl and Maier [74] and Ishida et al. [75] obtained similar results for plastid-targeted cryptomonad and raphidophyte proteins respectively. Interestingly, when the signal peptide is removed from the presequence, both diatom and cryptomonad precursor proteins are imported into pea chloroplasts in vitro [76,74]. This suggests that plastid-targeting presequences in these systems are bipartite, composed of a signal peptide domain and a region functionally homologous to a transit peptide.

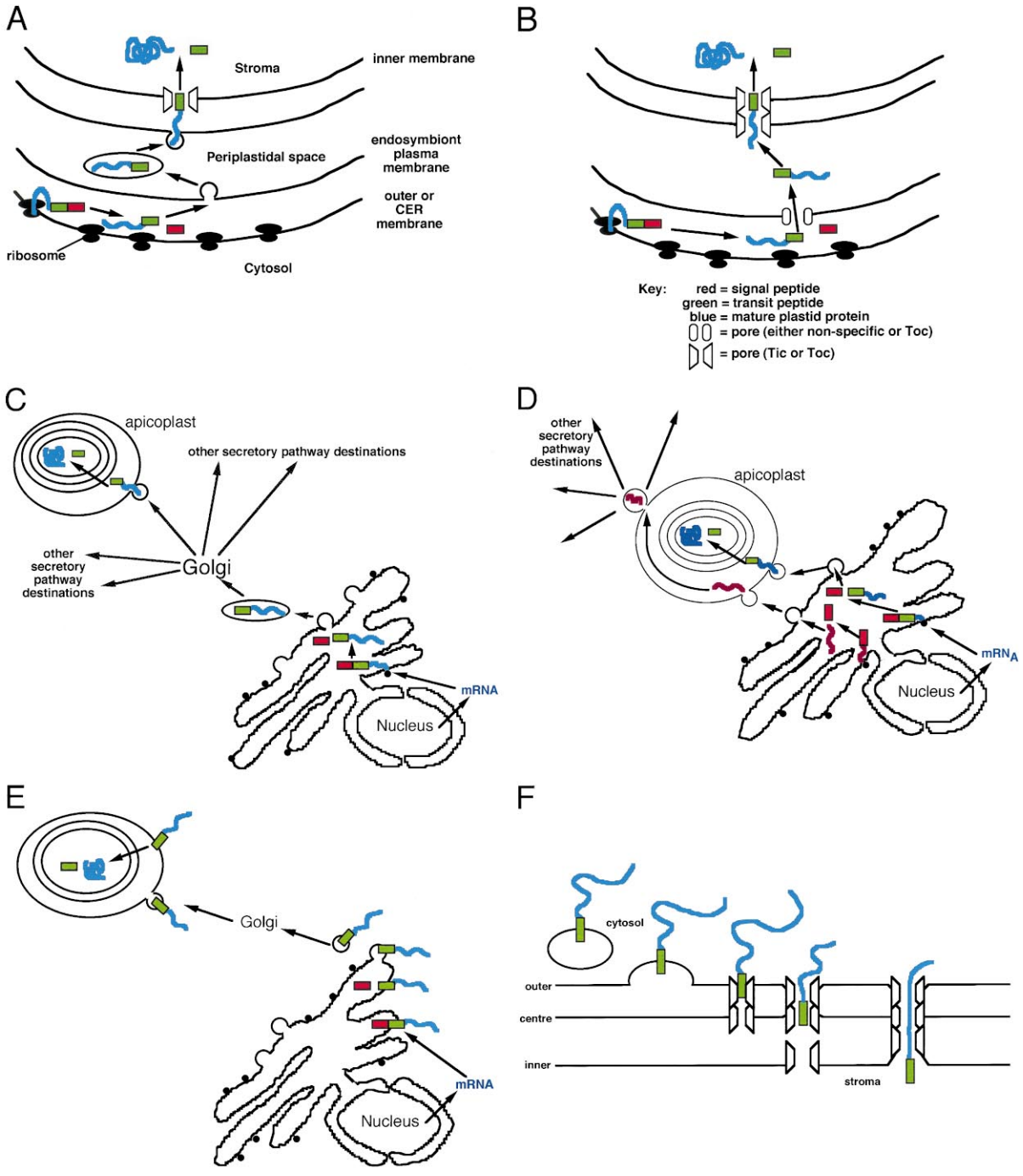
The discovery of the N-terminal signal peptide supported Gibbs' hypothesis [6] that plastid-targeted

Fig. 2. (A) Gibbs' model [6] of protein import into plastids containing a CER. The signal peptide of the presequence mediates co-translational import of proteins at ribosomes located on the CER membrane, with the signal peptide cleaved soon after import. Transport across the central two membranes involves vesicle shuttling, with proteins translocating through the innermost membrane via a protein complex. Once inside the stroma of the plastid, the transit peptide of the presequence is cleaved. (B) Kroth and Strottman's alternative model [80] of import into plastids containing a CER. It differs from Gibbs' model in that it proposes that protein translocation across the central two membranes is not vesicle mediated. Instead, protein complexes mediate translocation across the inner two membranes. Kroth and Strottman suggest a non-specific pore enables translocation across the remnant plasma membrane of the endosymbiont. An alternative model, initially proposed by Cavalier-Smith [41] and refined by van Dooren et al. [81], proposes this pore is a specific protein complex, possibly a duplicate of the Toc apparatus. (C) A model for ER/Golgi-mediated protein traffic to the plastids of apicomplexa. mRNA is transcribed at ribosomes (black circles) of the ER, resulting in signal peptide (red)-mediated translocation of apicoplast-directed proteins into the ER. Vesicle transport then brings proteins to the Golgi, from where they are sorted from other secretory pathway proteins and directed to their correct destination. Translocation across the outermost apicoplast membrane is achieved by fusion of these post-Golgi vesicles, which deposit their cargo into the space between the outer two membranes. Transit peptides (green) then mediate translocation across the remaining membranes. (D) A model for protein traffic to apicoplasts, where the apicoplast is located at start of the secretory pathway. After fusion of ER-derived vesicles carrying secretory pathway proteins with the outermost apicoplast membrane, proteins containing a transit peptide are recognized by protein complexes on the second outermost membrane, and translocated into the organelle. Secretory proteins lacking a transit peptide (purple) wash past the apicoplast and are taken up by vesicles that bud from the outermost membrane. From here, these vesicles carry proteins on to other secretory pathway destinations. (E) A working model for transport of nuclear-encoded plastid proteins to their destination in *Euglena*. A hydrophobic stop transfer signal (see Fig. 3) in the presequence (green) of plastid-targeted proteins prevents the protein from fully translocating into the ER. The protein thus travels as an integral membrane protein for the remainder of its Golgi-directed journey to the plastid, the C-terminal part of the protein (blue) remaining in the cytosol. After fusion with the outer plastid membrane, proteins are translocated into the stroma (see F). (F) Detailed model of protein translocation across the three membranes surrounding the plastid of *Euglena*, after [103]. Fusion of plastid precursor-containing vesicles with the outermost membrane results in the precursor being anchored by the stop transfer sequence in the membrane. Here it associates with protein complexes in this membrane. The 60-amino-acid long transit peptide, located on the lumenal side of the membrane, may then interact with a protein complex (possibly Toc-like) in the intermediate membrane, and then a protein complex in the inner membrane (possibly Tic-like). This results in the protein complexes of each of the membranes forming contact sites, and producing a translocation channel, through which proteins can migrate from the cytosol into the stroma.

→

proteins are translated on CER ribosomes and co-translationally translocated across the outer of the four plastid membranes. However, it is not known whether these proteins are also translated at ribosomes of the standard ER. In most heterokonts and cryptomonads, the CER is continuous with the ER and nuclear envelope (review in [39]). It appears that the CER is a functional part of the ER, with mastigonemes (flagellar hairs) synthesized on ribosomes and transported through the CER as well as the ER (review in [39]). Thus, it is possible that proteins translocated into the regular ER could be transported to the plastid, and conversely that secretory pathway proteins wash over the plastid before vesicle-mediated trafficking to other regions of the cell.

Early studies failed to observe ER/CER connections in some heterokonts, namely Raphidophyceae and some Eustigmatophyceae [39]. Interestingly, mastigonemes of the Raphidophyceae are only found in the ER. [39]. It was presumed that in this case (and possibly in other heterokonts and cryptomonads as well) a mechanism existed to direct the translation of plastid proteins to the CER [77]. Schwartzbach et al. [77] suggest this sorting occurs by plastid proteins having a plastid-specific signal peptide, which directs the protein to the CER. Since plastid-targeted proteins can be imported into canine microsomes, any differences in ER and CER signal peptides must be very subtle indeed. Recently, however, Ishida et al. [75] have observed ER/CER connections in the raphidophyte *Heterosigma akashiwo*. Interestingly, the CER of this organism contains few ribosomes. Ishida et al. suggest that all plastid-targeted proteins are co-translationally transported into the ER and then move to the plastid through the narrow ER/CER lumenal connections [75]. Furthermore, they argue that ER-targeting of plastid proteins is common in heterokont algae generally [75]. This raises the question of how chloroplast-directed pro-



teins are prevented from following the default secretory pathway to outside the cell. Presumably either some mechanism exists to prevent plastid-targeted proteins exiting the ER, or a mechanism is in place to retrieve plastid-targeted proteins from the Golgi apparatus. The development of genetic transformation in diatoms [78,79] should enable any differences that might exist between ER and CER signal peptides in diatoms to be determined. Transformation could also be used to determine whether the mecha-

nism for CER targeting is ER retention or Golgi retrieval, and to identify the sorting signal.

So how is transport across the remaining three membranes mediated? The first theory was suggested by Gibbs [6], who noted the existence of numerous vesicles in the periplastidal space of heterokonts. She hypothesized that after crossing the outermost membrane, vesicles carrying plastid-directed proteins bud from the second outermost membrane (thought to derive from the endosymbiont's plasma membrane; Fig. 2A). These vesicles were then envisaged to fuse with the subtending membrane (originally the outer membrane of a simple chloroplast) to deposit the cargo into the space between the inner pair of membranes (Fig. 2A). From here, proteins would be translocated across the inner membrane, perhaps using a translocation complex homologous to the Tic complex [53,54] of simple chloroplasts (Fig. 2A; [80]). An alternative mechanism has been proposed by Kroth and Strottman [80], whereby the plastid-targeted protein can freely diffuse through large pores or through a non-specific protein transporter in the third membrane. The presequence transit peptide would then interact with Toc and Tic homologues in the two innermost membranes and be imported into the plastid (Fig. 2B). Given the presence of ribosomes and other cellular material in the periplastid space of cryptomonads, a large pore in the vestigial plasma membrane of the endosymbiont would appear unlikely, as this would cause leakage of contents from the periplastidal space into the lumen of the ER. A refined version of the non-specific pore hypothesis was recently proposed [81], in which a duplicated Toc apparatus is located in the vestigial plasma membrane of the endosymbiont (Fig. 2B). This hypothetical apparatus is proposed to identify and remove transit peptide-containing proteins from the secretory pathway, diverting proteins across the second outermost membrane and into the periplastidal space [81]. Although translocation across the innermost membrane(s) undoubtedly requires protein complexes that interact with the transit peptide-like presequence domain, these complexes remain unidentified. Interestingly, however, a homologue of the Tic22 component of the inner membrane translocation complex in plants is present in the cryptomonad nucleomorph [17]. Once the protein has entered the plastid, the transit peptide would be cleaved by a stromal peptidase.

In cryptomonads, nuclear-encoded proteins must be translocated across four envelope membranes while nucleomorph-encoded proteins must only be translocated across the two innermost membranes. Plastid-targeted nucleomorph proteins contain N-terminal extensions that are not recognised by chloroplast transit peptide prediction programmes [74], whereas the nuclear-encoded proteins have a bipartite structure, containing both signal and transit peptide-like domains. Wastl and Maier [74] isolated cryptomonad plastids with the inner two membranes intact. Using this homologous system, they showed that a nucleomorph-encoded protein was imported into the plastid. They also made protein constructs of nuclear-encoded plastid proteins whereby they deleted the signal peptide region. Surprisingly, they found that this transit peptide-bearing protein was not imported into the cryptomonad plastid. Nevertheless, they found that both the nuclear-encoded and nucleomorph-encoded proteins were transported into pea chloroplasts, demonstrating that the N-terminal extension of the nucleomorph-encoded protein functionally resembles plant transit peptides. These results suggest that there might be two separate pathways into cryptomonad plastids, both of which are mediated by a transit peptide-like domain. One involves targeting nuclear-encoded proteins across the three innermost membranes surrounding the plastid, and another involves targeting nucleomorph-encoded proteins across the two innermost membranes. Nevertheless, this hypothesis is based on a single negative result [74] and needs further investigation. Comparisons of the transit peptide-like domains of nuclear-encoded plastid proteins to their nucleomorph counterparts would be a good place to start. A potential third import pathway into cryptomonad plastids was recently identified by Douglas et al. [17]. They found that DNA polymerase proteins necessary for replication of the nucleomorph were not encoded by nucleomorph genes [17]. Presumably, these genes are encoded in the nucleus, and must be transported from their site of translation across the outer two membranes only of the plastid, and into the periplastid space. Whether this import pathway has any similarities to the import of nuclear-encoded proteins

destined for the stroma or thylakoids of the plastid remains to be determined.

6. Protein import into plastids surrounded by four membranes, but lacking a CER

In diatoms and cryptomonads, an N-terminal signal peptide initiates translocation of plastid-targeted proteins across the CER, thus traversing the outermost plastid membrane. Although bound by four membranes, the plastids of apicomplexa (apicoplasts) and chlorarachniophytes appear to lack ribosomes on their outer membrane. No connections have ever been observed between the ER and the plastids of either apicomplexa or chlorarachniophytes. Nuclear-encoded, plastid-targeted genes have been identified in both apicomplexa and chlorarachniophytes [32–35,71,82,83]. Much like plastid-targeted proteins in heterokonts and cryptomonads, these proteins have a bipartite presequence comprising a signal peptide and a stromal targeting, transit peptide-like domain. No study has yet been conducted on protein targeting in chlorarachniophytes, so the remainder of this discussion focuses predominantly on apicomplexa.

Researchers of apicomplexa have the distinct advantage that their subjects are amenable to genetic manipulation. Waller et al. [32,34] and DeRocher et al. [82] constructed fusion proteins, whereby they fused the bipartite presequence of apicoplast-targeted fatty acid biosynthesis protein to the Green Fluorescent reporter Protein (GFP), and transformed cells of *T. gondii* and *P. falciparum*, both members of the apicomplexan phylum. They found the GFP fusion proteins were targeted to the apicoplast, indicating that the presequence alone is sufficient to target a protein to apicoplasts. Upon removal of the presequence signal peptide domain, the GFP fusion protein accumulated in the cell cytosol (and in one case the mitochondrion), while removal of the presequence transit peptide domain resulted in GFP secretion [34,82]. Secretion of GFP upon removal of the transit peptide suggests that the first step in apicoplast targeting is the co-translational transport of the protein into the ER. In the absence of further targeting signals, proteins in the secretory pathway are secreted from cells via the so-called default pathway. This suggests that the transit peptide-like domain functions in sorting proteins for apicoplast targeting within the secretory system.

DeRocher et al. [82] showed that the presequence transit peptide domain of *T. gondii* plastid proteins directed the in vitro import of GFP into pea chloroplasts. In a complementary experiment, Crawford and Roos (personal communication), constructed an 'artificial' bipartite plastid-targeting presequence containing a *T. gondii* signal peptide domain and an *Arabidopsis* transit peptide fused to GFP. This fusion protein was targeted to the apicoplast of *T. gondii*. These results confirm that the first step in apicoplast targeting is the signal peptide-mediated co-translational transport of the precursor into the ER. They also suggest that the function of the transit peptide-like domain is similar to plant transit peptides, possibly utilizing import machinery with similarities to the transit peptide-dependent machinery of simple chloroplasts. Interestingly, several putative homologues of the Toc and Tic complexes have been found in *P. falciparum*, with one (Tic22) shown to target to the apicoplast (C. Tonkin, G. McFadden, unpublished).

The major difference between protein targeting to apicomplexan plastids (and possibly chlorarachniophytes; [84]) and those of heterokonts and cryptomonads, is that proteins must be translocated into the ER, and then be sorted into transport vesicles that travel to and fuse with the outermost apicoplast membrane. In apicomplexa, this sorting is not a simple matter, with the secretory pathway targeting proteins to perhaps a dozen different destinations [81]. One can envisage two scenarios for traffic directed to the apicoplast. First, the apicoplast may be proximal to any divergence point in the endomembrane system (Fig. 2D). In this scenario all proteins entering the endomembrane system would pass by the apicoplast and those bearing transit peptides would be diverted into the apicoplast [81]. In this scenario no special information for vesicular traffic is required within the transit peptide component of the leader (Fig. 2D). Van Dooren et al. [81] hypothesized that this sorting may occur through interaction of the transit peptide with a duplicated set of Toc proteins located in the envelope membrane believed to be homologous to the plasma membrane of the endosymbiont (second outermost membrane). Proteins lacking a transit pep-

tide wash over the apicoplast and are packaged in vesicles that bud from the outer envelope membrane for transport to further destinations in the secretory pathway [81].

The second scenario is that the apicoplast is at the end of a specific branch in the endomembrane flow (Fig. 2C). For this to be true, the transit peptide would need to have an embedded signal (which must be recognized by factors on the surface of transport vesicles) that mediates accurate targeting (Fig. 2C). There is little evidence to discriminate between the two options at this stage and it is unknown whether apicoplast-destined proteins go through the Golgi [81]. One piece of indirect evidence perhaps supports the first scenario. When a plant transit peptide was substituted for a *Toxoplasma* transit peptide, targeting to the apicoplast was still efficient (Crawford and Roos, personal communication). Because this foreign transit peptide could mediate accurate targeting, it would seem unlikely that a specific vesicle-targeting domain occurs in the transit peptide-like region of the presequence, which makes a specific vesicle destination unlikely. However, this scenario predicts that vesicles bud from the apicoplast, something that has not been observed in these cells. A recent paper found that antibodies against a *P. falciparum* homologue to *N*-ethylmaleimide-sensitive factor, a protein which plays a crucial role in vesicle docking and fusion in eukaryotes (see review in [85]), localized this protein to regions surrounding the apicoplast [86]. This suggests that vesicles do fuse with the outer apicoplast membrane. Determining whether these vesicles contain only apicoplast proteins or whether they contain other secretory proteins might be a way of distinguishing between the two scenarios. Clearly, the question of how apicoplast proteins get from the ER to the plastid is an area that warrants further study.

Once the protein has been transported into the intermembrane space between the outer two apicoplast membranes, there are two conceivable mechanisms for transporting proteins across the three remaining membranes. Proteins may cross the space between the middle pair of membranes by vesicle budding in a manner similar to that proposed for heterokonts by Gibbs [6] (compare to Fig. 2A). Vesicles have not been observed in the 'periplastid' space of apicomplexa or in chlorarachniophytes [28], suggesting that vesicular transport may not be used for protein transport between the middle two envelope membranes in organisms lacking a CER. The second possibility is that proteins cross the central two membranes via protein complexes (compare to Fig. 2B). Tenuous homologues of the Toc complex have been found in *P. falciparum* (C. Tonkin, G. McFadden, unpublished), suggesting a Toc-like complex may play a role. However, the existence of such a Toc-like complex still requires experimental confirmation. An apicoplast-targeted homologue of Tic22, a protein believed to have a chaperone-like function [87], has been identified in *P. falciparum* and shown to target to the apicoplast (C. Tonkin, G. McFadden, unpublished). This suggests that in either scenario, a Tic-like complex may mediate translocation across the innermost membrane.

Once the protein has been transported across all four apicoplast envelope membranes, the transit peptide is cleaved ([35,32,34]; G. van Dooren, V. Su, G. McFadden, unpublished). A homologue of the stromal processing peptidase (SPP) of plants has recently been identified in *P. falciparum* (G. van Dooren, G. McFadden, unpublished). Although it remains to be determined if this protein is localized to the apicoplast, its presence adds weight to the suggestion that the eukaryotic endosymbiont that led to the apicomplexan line had already evolved several features of the transit peptide-dependent chloroplast import system seen in plants.

In summary, the targeting of nuclear-encoded proteins to the plastids of apicomplexa (and probably chlorarachniophytes) is mediated by a bipartite, signal peptide-containing presequence. However, unlike the equivalent process in CER-containing organisms, this targeting occurs via the general secretory pathway, with proteins co-translationally inserted into the ER and possibly transported in vesicles via the Golgi apparatus to the plastid. Reflecting this lengthened pathway, pulse-chase experiments suggest that in *P. falciparum* sorting and transport of proteins to apicoplasts is a slow process, requiring at least 45 min (V. Su, G. van Dooren, G. McFadden, unpublished).

7. Protein import into plastids surrounded by three membranes

Euglena and dinoflagellate plastids differ from the other organisms discussed in this review in that they are enclosed by three membranes: an inner, intermediate and outer membrane [5,88]. Ribosomes are absent from the outermost membrane of both dinoflagellate and euglenoid plastids. Therefore co-translational transport of chloroplast precursor proteins across this membrane does not commence at the plastid. Rather, similar to apicomplexa, it appears that transport occurs through the endomembrane system upstream of the plastid. Though morphologically similar, little work has examined protein targeting to peridinin-containing dinoflagellate plastids, and this review will focus predominantly on *Euglena* plastid import.

Just as electron microscopy first led Gibbs to propose co-translational precursor protein transport across the outermost membrane of organisms with CER, immunogold electron microscopy provided the first definitive evidence for involvement of the secretory pathway in transport of proteins from the cytoplasm to the *Euglena* chloroplast. Using immunoelectron microscopy to study light-harvesting chlorophyll *a*/*b*-binding protein of Photosystem II (LHCPII) induction upon transfer of *Euglena* from low to high light, Osafune et al. [89] discovered that LHCPII was present in the Golgi apparatus. In *Euglena* LHCPII synthesis and accumulation can be induced upon exposing dark-grown (non-green) cells to light [90]. Interestingly, 12 h after inducing LHCPII synthesis, the protein is detected only in the Golgi. After 24 h of light exposure, the time of maximum LHCPII synthesis, both the Golgi apparatus and chloroplast are immunogold labelled [91]. By 72 h of light exposure, when LHCPII synthesis ceases [90], only the chloroplast is immunogold labelled, with LHCPII not detected in the Golgi [91].

In addition to localizing LHCPII to the Golgi apparatus and chloroplast, immunogold electron microscopy localized LHCPII to a structure known as the compartmentalized osmiophilic body (COS) [91–93]. The COS has an osmiophilic core with reticulate extensions of the same material interconnecting adjacent core regions. Serial sections indicate the COS is composed of tightly packed membranous structures that appear connected to the ER (T. Osafune, unpublished). After exposure of *Euglena* cells to light, LHCPII was detected in the COS several hours before the Golgi [91,92]. This suggests that the COS may either be a specialized region of the ER where chloroplast proteins are synthesized and packaged for transport to the Golgi apparatus, or an intermediate compartment between the ER and Golgi apparatus.

In vivo pulse-chase subcellular fractionation studies have provided direct proof that plastid-targeted *Euglena* proteins are transported via the Golgi-to-chloroplast secretory pathway [94,95]. Sulli and Schwartzbach [94,95] pulsed cells with [^{35}S]sulfate and fractionated cellular compartments on a sucrose gradient. At the end of a pulse, the precursor of SSU (pSSU) and pLHCPII are found in the ER and Golgi apparatus [90,95]. The localization of these newly synthesized proteins in the ER is consistent with their co-translational import at membrane-bound polysomes [96]. During the chase, the fraction of pLHCPII and pSSU in the ER decreases and the amount in the Golgi apparatus increases. This increase is associated with the appearance of mature SSU and LHCPII in the chloroplast [90,95]. Interestingly, pSSU and pLHCPII are both synthesized as polyprotein precursors composed of 8 mature proteins covalently linked by a conserved decapeptide [97,98]. Dinoflagellates are the only organisms other than *Euglena* where chloroplast polyprotein precursors have been reported [99,100]. What, if any, relationship there is between three chloroplast envelope membranes and the occurrence of polyprotein precursors remains is unknown.

Having established the route plastid-directed proteins take to their destination, Sulli and Schwartzbach made the unexpected discovery that pulse-labelled pLHCPII and pSSU could not be removed from ER and Golgi membrane fractions by washing with Na_2CO_3, a treatment that removes peripheral but not integral membrane proteins. This suggests that plastid-targeted proteins travel through the secretory pathway as integral membrane proteins [94,95]. Both pLHCPII and pSSU were partially digested by trypsin treatment of intact ER and Golgi membranes, indicating that portions of the protein remain on the cytoplasmic membrane face. Taken together, these results suggest that *Euglena* chloro-

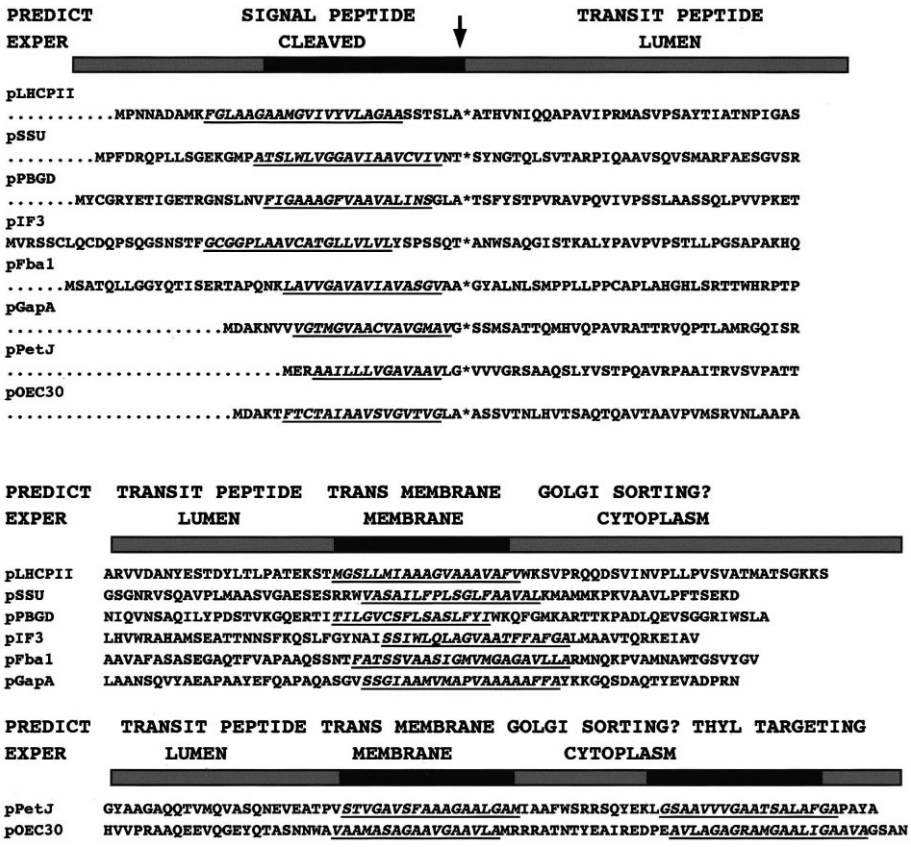


Fig. 3. Comparison of the deduced amino acid sequence of the N-terminal presequence of the *Euglena* light-harvesting chlorophyll *a*/*b*-binding protein (pLHCPII [102]), small subunit of ribulose-bis-phosphate carboxylase (pSSU [98]), porphobilinogen deaminase (pPBGD [115]), chloroplast initiation factor 3(pIF3 [116]), fructose-1,6-bisphosphate aldolase (pFba1 [117]), chloroplast glyceraldehyde-3-phosphate dehydrogenase (pGapA [118]), the thylakoid (THYL) lumenal protein cytochrome c6 (pPetJ [119]) and the thylakoid lumenal protein the 30 kDa protein of the oxygen-evolving complex (pOEC30 [119]). Computer analysis was used to predict (PREDICT) signal peptides, transit peptides, thylakoid targeting signals and transmembrane domains [101,104]. Experimentally (EXPER) determined domain functions are indicated (see text). Sequences were aligned at the computer-predicted signal peptidase cleavage site indicated by an arrow and asterisk in the sequences. Hydrophobic domains are indicated by dark bars, and are italicized and underlined in the sequence.

plast proteins are transported in vesicles via the ER and Golgi apparatus to the chloroplast as integral membrane proteins, oriented with a major portion of the protein located on the cytoplasmic side of the vesicle.

All *Euglena* chloroplast protein presequences analysed to date have a common domain structure (Fig. 3). Like the presequences of plastid-targeted proteins in the other organisms discussed in this review, *Euglena* precursors contain a hydrophobic N-terminal domain identified by computer algorithms as a signal peptide [101]. However, the region following this signal peptide seems somewhat unusual. If the presequences are aligned at the computer-predicted site of signal peptidase cleavage, a second hydrophobic domain occurs in the same relative position in all of the presequences; approximately 60 amino acids from the predicted signal peptidase cleavage site (Fig. 3). Between these two hydrophobic zones lies a transit peptide-like domain (Fig. 3; see shortly).

In vitro canine microsome studies have shed some initial light on the function of the conserved, second hydrophobic region. The N-terminal 35 amino acids of the *Euglena* pLHCPII presequence initiates the co-translational insertion of protein into microsomal membranes, with the presequence being cleaved [102]. This confirms that the first hydrophobic domain functions as a signal peptide. However,

pLHCPII is not translocated into the microsomal lumen but becomes an integral membrane protein. Further studies have demonstrated that pLHCPII is oriented so that the N-terminus is in the lumen and the C-terminus of the protein remains outside the microsomal membrane [103], topologically corresponding to the cytosol. Deletion studies indicated that the second hydrophobic domain of the pLHCPII presequence (Fig. 3) functions as the stop transfer membrane anchor sequence.

Computer analysis [104] predicts the presequence portion following the first hydrophobic region to be a stromal-targeting transit peptide (Fig. 3), similar to that which directs import into plant chloroplasts. This suggests that *Euglena* chloroplast proteins may have a similar mechanism to other plastids for crossing the inner two chloroplast membranes. In vitro import studies with isolated *Euglena* and pea chloroplasts have provided clues as to the function of this transit peptide-like region. pSSU was not directly imported into *Euglena* chloroplasts, even when the signal peptide presequence domain was deleted (Vacula and Schwartzbach, unpublished). However, incubation of a [^{35}S]pLHCPII-containing Golgi membrane fraction with isolated *Euglena* chloroplasts does result in the light-, ATP-, and GTP-dependent import of pLHCPII into the chloroplast (Slavikova and Schwartzbach, unpublished). These results suggest that proteins can only cross the outer chloroplast membrane via the secretory pathway. Furthermore, pSSU was only imported into peas chloroplasts when the first hydrophobic domain was deleted (Vacula and Schwartzbach, unpublished), confirming computer predictions that the region following the signal peptide is a functionally conserved transit peptide. Studies with presequence deletion mutants identified the stromal-targeting transit peptide domain as the region between the signal peptidase cleavage site and the start of the hydrophobic membrane anchor domain (Fig. 3; Vacula and Schwartzbach, unpublished). This region is predicted to be in the lumen of the transport vesicle, and given its homology to plant transit peptides may function in a similar manner to cross the inner two chloroplast membranes. Upon fusion of the transport vesicle with the outermost envelope membrane, the transit peptide may interact with an intermediate and inner membrane import apparatus evolutionarily related to the Toc and Tic import proteins of plant chloroplasts (Fig. 2F).

In vitro studies of the nuclear-encoded plastid protein porphobilinogen deaminase (PBGD) found that pPBGD was not imported into canine microsomes [105]. Nevertheless, the signal peptide domain of pPBGD was cleaved by the *Escherichia coli* leader peptidase and cyanobacterial processing peptidase, indicating it has some properties of a signal peptide [105]. pSSU is also not processed during co-translational import into canine microsomes even though in vivo studies clearly demonstrate its transport from the ER-Golgi-chloroplast import route [95]. Furthermore, the presequence of pPBGD is similar to those of other plastid-targeted proteins in *Euglena* (Fig. 3). These observations suggest that pPBGD may in fact follow the standard ER/Golgi mediated plastid import pathway. As a result, the in vitro functionality of the pPBGD signal peptide domain as an ER-targeting sequence remains an open question.

The working model that has emerged for *Euglena* chloroplast protein import holds that proteins are transported via the secretory pathway. Precursor proteins are co-translationally inserted into the ER then transported to the Golgi apparatus as integral membrane proteins [103]. These proteins are oriented with an approximately 60-residue-long transit peptide-like domain in the lumen and the remainder of the presequence and mature protein in the cytoplasm (Fig. 2E). In the Golgi apparatus, the proteins must be sorted and packaged into transport vesicles destined for the chloroplast. Nothing is known about how this sorting and packaging occurs. It is tempting to speculate that the cytoplasmic domain of the presequence plays a role in assembly of the protein coat on the membrane surface that is needed for vesicle formation. The Golgi vesicles carrying membrane-bound precursors apparently fuse with the outermost plastid membrane (Fig. 4), and what occurs next can be guessed at.

The precursor is anchored in the outermost envelope membrane with the 60-amino-acid long transit peptide domain in the intermembrane space. It is feasible that the transit peptide interacts with a putative intermediate membrane import complex, which may be related to the Toc complex of higher plants (Fig. 2F). This interaction could produce a conformational change in the presequence, allowing

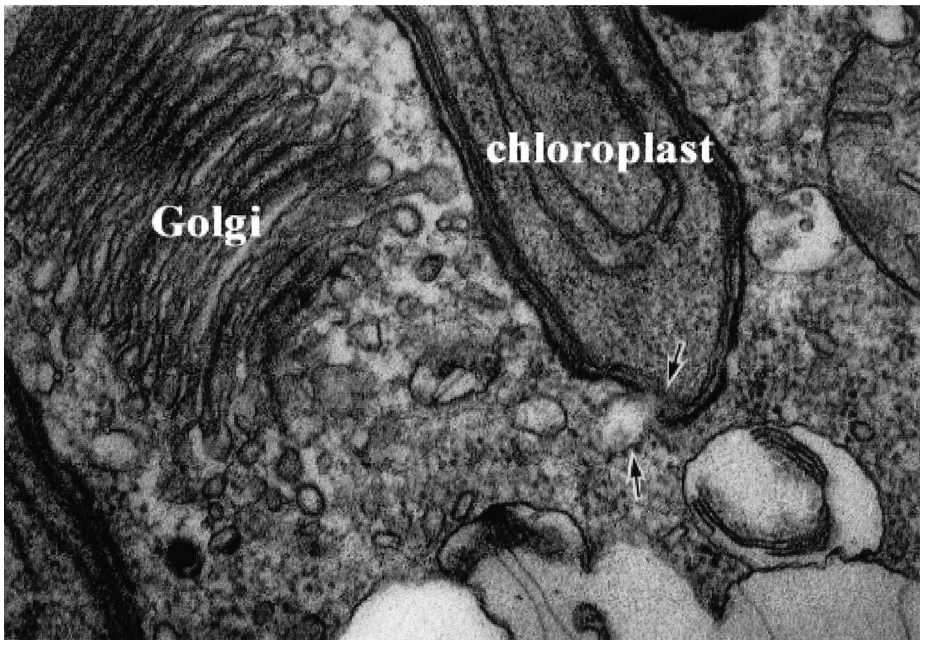


Fig. 4. Electron microscopy image of *Euglena*, showing a Golgi-derived vesicle fusing with the outermost membrane of the *Euglena* chloroplast (see arrows). Plastid-targeted proteins are present as integral membrane proteins in the vesicle, and this process results in their localization to the outer chloroplast membrane, from where they cross the remaining membranes into the chloroplast stroma (see text for details).

the integral membrane region to migrate out of its lipid environment into a proteinaceous outer membrane translocation channel, a process analogous, but reversed, to the movement of a protein from the ER translocation channel into the membrane [106]. Interaction between the outer and intermediate membrane translocation channels could then form a contact site, facilitating transit peptide translocation across these outer two membranes. After moving through the intermediate membrane, the transit peptide could interact with an inner membrane import complex, which may be related to the Tic complex of higher plants. Freeze fracture electron microscopy reveals regions where the *Euglena* intermediate membrane is in contact with the inner membrane [107]. These contact sites are reminiscent of the contact sites formed by the translocation machinery of the outer and inner envelope membranes of simple chloroplasts [53,54], suggesting that they form a channel for protein transport across the intermediate and inner membranes. Formation of such a channel would allow proteins to translocate across the three membranes and into the stroma, where the presequence is cleaved (Fig. 2F). Thylakoid lumenal proteins have an additional hydrophobic domain near the presequence C-terminus (Fig. 3) that computer programs [101] identify as the hydrophobic core of a signal peptide. This could function, as it does in higher plants, as a thylakoid targeting domain.

In summary, like protein targeting to other complex plastids, plastid targeting in *Euglena* occurs via the secretory system. However, this targeting is mediated by a tripartite presequence (Fig. 3). The first part of this sequence functions as a signal peptide to direct proteins to the ER. The second region appears homologous to plant transit peptides, and may function in directing proteins across the inner membranes of the plastid. The unique third domain is a hydrophobic stop transfer signal, which ensures *Euglena* precursors are transported to the plastid as integral membrane proteins [94,95]. This hydrophobic membrane anchor domain has no obvious parallel in the other groups containing secondary plastids. So, why did *Euglena* presequences contain such a stop transfer membrane anchor system, when in other complex plastids a signal peptide-transit peptide bipartite system seems sufficient? One possibility is that it evolved to deal with the transport of transmembrane proteins through the secretory system. Studies of *Euglena* LHCPII have shown that when

the stop transfer domain of the presequence is removed, hydrophobic transmembrane domains within the mature protein prevent the protein fully translocating into microsomal membranes [103], behaving much like the stop transfer region in embedding the protein into the membrane. During the evolution of a plastid-targeting protein mechanism, hydrophobic transmembrane domains such as that in LHCPII may have prevented complete translocation of these proteins, jamming them into the ER membrane instead. Rather than evolve a way of getting these proteins into the ER/secretory pathway, it may have been sufficient for *Euglena*'s ancestors to import all nuclear-encoded chloroplast proteins with the mature part of the protein remaining in the cytoplasm. This required the evolution of a stop transfer membrane anchor sequence in all *Euglena* presequences, and a mechanism to transfer integral membrane proteins from the outer chloroplast envelope into an import channel. In all other organisms with complex plastids, the limited evidence suggests precursors are translocated completely through the ER/CER membrane into the endomembrane lumen. This includes the LHCPII homologues of these systems, all of which, like *Euglena*, have three transmembrane helices [40,73]. As more information is obtained for import into other complex plastids, comparisons with the unique *Euglena* system may provide further insights as to how it may have evolved.

Studies of protein targeting to the peridinin-containing plastids of dinoflagellates (the only other group with three membranes bounding their plastids) is still in its infancy. The presequences of several nucleus-encoded dinoflagellate plastid proteins are available (references in [108]), but no targeting studies have been reported. Most, but apparently not all, the dinoflagellate presequences carry a putative signal peptide and a downstream region that is hydrophilic and perhaps equivalent to the transit peptide-like domain of other presequences. The signal peptide domain at the N-terminus of the dinoflagellate presequences suggests they are transported via the secretory pathway. Given that dinoflagellates do not appear to be related to euglenoids, there is no particular reason to suspect they have a similar integral membrane-based targeting mechanism. Indeed, with recent studies indicating dinoflagellate plastids are related to those of apicomplexa [44,48,49], it will be intriguing to see how protein trafficking in these two systems is related, especially given dinoflagellates have one fewer membrane than apicomplexa. Needless to say, studies of chloroplast targeting in dinoflagellates are eagerly awaited.

A final intriguing factor about protein targeting in dinoflagellates concerns those organisms containing tertiary plastids. Many dinoflagellates have lost their peridinin-containing plastids and become heterotrophic. Several of these dinoflagellates have since acquired plastids by endosymbiosis of organisms containing secondary plastids, including from haptophyte algae, diatoms and cryptomonads [47,112,113]. No study has yet determined whether proteins are targeted from the host cell nucleus to within these plastids, and if so how this process occurs. However, it is interesting to note that dinoflagellates containing haptophyte-derived plastids, contrary to what might be expected, seem to be surrounded by fewer than six membranes [112].

8. Conclusions and evolutionary considerations

Complex plastids are believed to have originated through secondary endosymbiosis, when a eukaryotic host acquired a phototrophic eukaryotic endosymbiont. The most convincing evidence for secondary endosymbiosis is the presence of remnant eukaryotic genomes, nucleomorphs, associated with the plastids of cryptomonads and chlorarachniophytes. These genomes, which originally contained most of the genes required for plastid function, are greatly reduced [17,114], and have been entirely lost in most organisms discussed here [47]. A consequence of secondary endosymbiosis was the massive transfer of genes from the symbiont to the host nucleus. Although there are multiple reasons why this gene transfer may have occurred [108], its occurrence presented a major problem for the host cells. With symbiont genes in their nucleus, these hosts needed to find some way of targeting plastid proteins back through the multiple plastid envelope membranes. Given that secondary endosymbiosis of plastids appears to have occurred on at least three separate occasions (but see [41]), this protein import system must have evolved independently at least three times. Remarkably, it

appears that in each case similar import systems evolved.

The plastid-targeted proteins of heterokonts, cryptomonads, apicomplexa dinoflagellates, chlorarachniophytes and euglenoids all appear to contain complex presequences composed of at least two elements, a proximal signal peptide element that mediates cotranslations import of these proteins into the ER, and a downstream transit peptide-like element. In apicomplexa, cryptomonads and heterokonts, experimental evidence suggests that this transit peptide-like domain is sufficient to mediate transport across the remaining membranes. The exception is in *Euglena*, where a hydrophobic, third domain is required for correct transport of proteins in secretory vesicles destined for the chloroplast. Transit peptide-like elements in *Euglena*, apicomplexa, cryptomonads and heterokonts can also mediate in vitro protein import into isolated 'simple' plastids of higher plants, suggesting that they contain essential elements of transit peptides.

This apparent parallel evolution of secretory-mediated transport to complex plastids provides a biochemical 'fossil record' of the evolution of these organisms. Six hundred million years or so ago, when the two endosymbiotic partners first met, they were separate, independent cells. At some point, the host began permanently engulfing or enslaving the symbiont, possibly incorporating the symbiont into a phagocytotic-type vacuole (which lies topologically 'outside' the host). Given that proteins can be targeted outside the cell via the secretory pathway, the simplest way to return proteins to the endosymbiont would be to add a signal peptide to the presequence. Thus, the ideal system for targeting nuclear-encoded plastid proteins to the endomembrane localized endosymbiont already existed. In a sense, getting proteins *to* the endocytotic vesicle encasing the photosynthetic symbiont was the easy part, requiring only a mechanism to ensure post-Golgi vesicles were directed to the correct compartment.

Once the protein reached the endomembrane lumen, a mechanism was required to cross the (presumably three) remaining membranes surrounding the symbiont. The first obstacle was the plasma membrane of the host. At this stage the protein is 'outside' both members of the partnership and we have little understanding of how this transport step might have been achieved. The problem of how to transport proteins across this membrane that faced host cells $\sim$600 million years ago is the exact same problem that faces researchers today. The insight of Gibbs 22 years ago, and the consequent work of several other groups, have shown us how the signal peptide can get proteins to the secretory system, but the processes by which proteins cross this third membrane remains elusive. This obstacle begs the question why have a membrane at all? We think the membrane was necessary initially to separate the endosymbiont cytoplasm from the lumen of the endomembrane system. In two algae, euglenoids and dinoflagellates, this membrane has apparently disappeared. Intriguingly, in euglenoids a different version of targeting appears to have evolved as compared to four membrane plastids. It will be of interest to learn how dinoflagellates target proteins.

Some progress has been made on the mechanism by which proteins are transported across the two innermost envelope membranes. Interestingly, this process may involve a mechanism that has evolved in parallel across the several evolutionary distinct groups, and again it seems that evolution has simply recycled existing mechanisms. The region of plastid-targeting presequences following the signal peptide has many of the features of a transit peptide, and in *Euglena*, cryptomonads, apicomplexa and diatoms has been shown capable of targeting proteins into plant plastids. This suggests that the mechanisms of protein import across the two innermost membranes may be similar to that in plants, and that the transit peptide import system of 'simple' chloroplasts evolved prior to the several secondary endosymbioses.

The evolution of protein targeting from host cells to secondary endosymbionts appears superficially similar in a range of organisms. Initially, proteins encoded by the nucleus of the symbiont (the consequent nucleomorph) contained transit peptides to target them across the two membranes surrounding the plastid. Indeed, proteins encoded by the cryptomonad nucleomorph appear to have kept this transit peptide-mediated system. After transfer of these transit peptide-containing genes from the symbiont 'nucleomorph' to the host nucleus, the addition of a signal peptide to the N-terminus was (generally) sufficient to target the resulting proteins across the

outermost and two innermost membranes of the plastid. Theoretically, then, plastid genes transferred from the nucleomorph to the nucleus utilized two existing transport systems and needed to invent only one new transport system to cross the plasma membrane of the endosymbiont [108]. Although unifying principles have emerged from studies of protein import into complex chloroplasts from diverse organisms, it is important to realize the each system is, often subtly, different.

To understand fully how secondary plastids evolved, it would be necessary to hop into a time machine, plankton filter in hand, and return to the days of trilobites. Indeed, we would probably need to go back to several points in time, as the evolution of secondary plastids probably took place over millions of years, and probably occurred numerous times. Sadly, we can only guess at what really happened. Nevertheless, thinking of protein targeting in an evolutionary framework is potentially worthwhile. Addressing questions such as why multiple membranes exist around complex plastids, or why dinoflagellate and apicomplexan plastids, which appear closely related, differ in their number of membranes, may provide clues as to how proteins cross these membranes.

Acknowledgements

We thank Chris Tonkin for critically reading the manuscript. Work in the authors' laboratories was supported by the United States National Science Foundation (MCB-96308 and MCB-0080345 to S.D.S.), the Ministry of Education, Science, Sports and Culture of Japan (Grant-in-Aid for Scientific Research (C) 09640795 to T.O.), the Japan Society for the Promotion of Science (T.O.), the Australian Research Council (G.I.M.), and the National Health and Medical Research Council (G.I.M.). G.I.M. is a Howard Hughes Medical Institute International Research Scholar. G.v.D. is supported by an Australian Postgraduate Award.

References

[1] T.M. Han, B. Runnegar, Science 257 (1992) 232–235.
[2] M.F. Liaud, C. Valentin, W. Martin, F.Y. Bouget, B. Kloareg, R. Cerff, J. Mol. Evol. 38 (1994) 319–327.
[3] W. Martin, B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, K. Kowallik, Nature 393 (1998) 162–165.
[4] D. Moreira, H. Le Guyader, H. Philippe, Nature 405 (2000) 69–72.
[5] S.P. Gibbs, Can. J. Bot. 56 (1978) 2883–2889.
[6] S.P. Gibbs, J. Cell Sci. 35 (1979) 253–266.
[7] J.M. Whatley, P. John, F.R. Whatley, Proc. R. Soc. Lond. B 204 (1979) 165–187.
[8] R.E. Lee, S. Afr. J. Sci. 73 (1977) 179–182.
[9] L.B. Zablen, M.S. Kissil, C.R. Woese, D.E. Buetow, Proc. Natl. Acad. Sci. USA 72 (1975) 2418–2422.
[10] T. Cavalier-Smith, Biol. J. Linn. Soc. 17 (1982) 289–306.
[11] M.A. Gillot, S.P. Gibbs, J. Phycol. 16 (1980) 558–568.
[12] G.I. McFadden, J. Cell Sci. 95 (1990) 303–308.
[13] A.D. Greenwood, H.B. Griffiths, U.J. Santore, Br. Phycol. J. 12 (1977) 119.
[14] P. Hansmann, H. Falk, P. Sitte, Z. Naturforsch. 40c (1985) 933–935.
[15] M. Ludwig, S.P. Gibbs, Protoplasma 127 (1985) 9–20.
[16] S. Zauner, M. Fraunholz, J. Wastl, S. Penny, M. Beaton, T. Cavalier-Smith, U.-G. Maier, S. Douglas, Proc. Natl. Acad. Sci. USA 97 (2000) 200–205.
[17] S. Douglas, S. Zauner, M. Fraunholz, M. Beaton, S. Penny, L.-T. Deng, X. Wu, M. Reith, T. Cavalier-Smith, U.-G. Maier, Nature 410 (2001) 1091–1096.
[18] S.E. Douglas, C.A. Murphy, D.F. Spencer, M.W. Gray, Nature 350 (1991) 148–151.
[19] U.-G. Maier, C.J.B. Hoffman, S. Eschbach, J. Wolters, G. Igloi, Mol. Gen. Genet. 230 (1991) 155–160.
[20] G.I. McFadden, P.R. Gilson, S.E. Douglas, J. Cell Sci. 107 (1994) 649–657.
[21] D.J. Hibberd, R.E. Norris, J. Phycol. 20 (1984) 310–330.
[22] Y. van de Peer, S.A. Rensing, U.-G. Maier, R. de Wachter, Proc. Natl. Acad. Sci. USA 93 (1996) 7732–7736.
[23] G. McFadden, P. Gilson, Trends Ecol. Evol. 10 (1995) 12–17.
[24] W.F. Martin, M. Müller, Nature 392 (1998) 37–41.
[25] D. Moreira, P. Lopez-Garcia, J. Mol. Evol. 47 (1998) 517–530.
[26] R.E. Lee, P. Kugrens, Protist 149 (1998) 341–345.
[27] R.E. Lee, P. Kugrens, Phycologia 39 (2000) 167–172.
[28] D.J. Hibberd, R.E. Norris, J. Phycol. 20 (1984) 310–330.
[29] R.J.M. Wilson, P.W. Denny, P.R. Preiser, K. Rangachari, K. Roberts, A. Roy, A. Whyte, M. Strath, D.J. Moore, P.W. Moore, D.H. Williamson, J. Mol. Biol. 261 (1996) 155–172.
[30] G.I. McFadden, M.E. Reith, J. Munholland, N. Langunnasch, Nature 381 (1996) 482.
[31] S. Kohler, C.F. Delwiche, P.W. Denny, L.G. Tilney, P. Webster, R.J.M. Wilson, J.D. Palmer, D.S. Roos, Science 275 (1997) 1485–1489.
[32] R.F. Waller, P.J. Keeling, R.G.K. Donald, B. Striepen, E. Handman, N. Langunnasch, A.F. Cowman, G.S. Besra, D.S. Roos, G.I. McFadden, Proc. Natl. Acad. Sci. USA 95 (1998) 12352–12357.

[33] H. Jomaa, J. Wiesner, S. Sanderbrand, B. Altincicek, C. Weidemeyer, M. Hintz, I. Turbachova, M. Eberl, J. Zeidler, H.K. Lichtenthaler, D. Soldati, E. Beck, Science 285 (1999) 1573–1576.
[34] R.F. Waller, M.B. Reed, A.F. Cowman, G.I. McFadden, EMBO J. 19 (2000) 1794–1802.
[35] N. Surolia, A. Surolia, Nat. Med. 7 (2001) 167–173.
[36] G.I. McFadden, D.S. Roos, Trends Microbiol. 6 (1999) 328–333.
[37] J.A.P. Diniz, E.O. Silva, R. Lainson, W. de Souza, Parasitol. Res. 86 (2000) 971–977.
[38] J. Hopkins, R. Fowler, S. Krishna, I. Wilson, G. Mitchell, L. Bannister, Protist 150 (1999) 283–295.
[39] S.P. Gibbs, Int. Rev. Cytol. 72 (1981) 49–99.
[40] D.G. Durnford, J.A. Deane, S. Tan, G.I. McFadden, E. Gantt, B.R. Green, J. Mol. Evol. 48 (1999) 59–68.
[41] T. Cavalier-Smith, J. Eukaryot. Microbiol. 47 (1999) 347–366.
[42] G.I. McFadden, R.F. Waller, Bioessays 19 (1997) 1033–1040.
[43] J. Blanchard, J.S. Hicks, J. Eukaryot. Microbiol. 46 (1999) 367–375.
[44] N.M. Fast, J. Kissinger, D. Roos, P. Keeling, Mol. Biol. Evol. 18 (2001) 418–426.
[45] M.E. Reith, Annu. Rev. Plant Physiol. Plant Mol. Biol. 46 (1995) 549–575.
[46] C.F. Delwiche, J.D. Palmer, Plant Syst. Evol. 11 (Suppl) (1997) 51–86.
[47] C.F. Delwiche, Am. Nat. 154 (1999) S164–S177.
[48] Z. Zhang, B. Green, T. Cavalier-Smith, J. Mol. Evol. 51 (2000) 26–40.
[49] G.I. McFadden, Genome Biol. 1 (2000) 1026.
[50] J.P. Bujak, G.L. Williams, Can. J. Bot. 59 (1981) 2077–2087.
[51] T. Cavalier-Smith, Trends Plant Sci. 5 (2000) 174–182.
[52] T. Cavalier-Smith, in: F. Round, D. Chapman (Eds.), Progress in Phycological Research, Vol. 3, Biopress, Bristol, 1986, pp. 309–347.
[53] D. Jackson-Constan, K. Keegstra, Plant Physiol. 125 (2001) 1567–1576.
[54] U. Vothknecht, J. Soll, Biol. Chem. 381 (2000) 887–897.
[55] J. Soll, R. Tien, Plant Mol. Biol. 38 (1998) 191–207.
[56] S. Richter, G. Lampa, J. Cell Biol. 147 (1999) 33–44.
[57] J. Jakowitsch, C. Neumann-Spallart, Y. Ma, J. Steiner, H.E.A. Schenk, H.J. Bohnert, W. Löffelhardt, FEBS Lett. 381 (1996) 153–155.
[58] K.E. Apt, N. Hoffman, A.R. Grossman, J. Biol. Chem. 268 (1993) 16208–16215.
[59] B. Bolter, T. May, J. Soll, FEBS Lett. 441 (1998) 59–62.
[60] G.I. McFadden, Curr. Opin. Plant Biol. 2 (1999) 513–519.
[61] S. Reumann, K. Keegstra, Trends Plant Sci. 4 (1999) 302–307.
[62] S. Reumann, J. Davila-Aponte, K. Keegstra, Proc. Natl. Acad. Sci. USA 96 (1999) 784–789.
[63] A. Grossman, A. Manodori, D. Snyder, Mol. Gen. Genet. 224 (1990) 91–100.
[64] P.G. Pancic, H. Strotmann, FEBS Lett. 320 (1993) 61–66.
[65] K.E. Apt, S.K. Clendennen, D.A. Powers, A.R. Grossman, Mol. Gen. Genet. 246 (1995) 455–464.
[66] C. Passaquet, C. Lichtl, Plant Mol. Biol. 29 (1995) 135–148.
[67] P.G. Kroth-Pancic, Plant Mol. Biol. 27 (1995) 825–828.
[68] L. Caron, D. Douady, M. Quinetszely, S. Degoer, C. Berkaloff, J. Mol. Evol. 43 (1996) 270–280.
[69] D.G. Durnford, R. Aebersold, B.R. Green, Mol. Gen. Genet. 253 (1996) 377–386.
[70] M.-F. Liaud, U. Brandt, M. Scherzinger, R. Cerff, J. Mol. Evol. 44 (Suppl 1) (1997) S28–37.
[71] J. Deane, M. Fraunholz, V. Su, U.-G. Maier, W. Martin, D. Durnford, G. McFadden, Protist 151 (2000) 239–252.
[72] K.E. Apt, D. Bhaya, A.R. Grossman, J. Appl. Phycol. 6 (1994) 225–230.
[73] D. Bhaya, A. Grossman, Mol. Gen. Genet. 229 (1991) 400–404.
[74] J. Wastl, U.-G. Maier, J. Biol. Chem. 275 (2000) 23194–23198.
[75] K. Ishida, T. Cavalier-Smith, B.R. Green, J. Phycol. 36 (2000) 1135–1144.
[76] M. Lang, K.E. Apt, P.G. Kroth, J. Biol. Chem. 273 (1998) 30973–30978.
[77] S.D. Schwartzbach, T. Osafune, W. Loffelhardt, Plant Mol. Biol. 38 (1998) 247–263.
[78] K.E. Apt, P.G. Kroth-Pancic, A. Grossman, Mol. Gen. Genet. 252 (1996) 572–579.
[79] L.A. Zaslavskaia, J.C. Lippmeier, P.G. Kroth, A.R. Grossman, K.E. Apt, J. Phycol. 36 (2000) 379–386.
[80] P. Kroth, H. Strotmann, Physiol. Plant 107 (1999) 136–141.
[81] G.G. van Dooren, R.F. Waller, K.A. Joiner, D.S. Roos, G.I. McFadden, Parasitol. Today 16 (2000) 421–427.
[82] A. DeRocher, C.B. Hagen, J.E. Froehlich, J.E. Feagin, M. Parsons, J. Cell Sci. 113 (2000) 3969–3977.
[83] M. Vollmer, N. Thomsen, S. Wiek, F. Seeber, J. Biol. Chem. 276 (2001) 5483–5490.
[84] A. Bodyl, Bot. Acta 110 (1997) 395–400.
[85] M.J. Clague, Biochem. J. 336 (1998) 271–282.
[86] M. Hayashi, S. Taniguchi, Y. Ishizuka, H.-S. Kim, Y. Wataya, A. Yamamoto, Y. Moriyama, J. Biol. Chem. 276 (2001) 15249–15255.
[87] A. Kouranov, X.J. Chen, B. Fuks, D.J. Schnell, J. Cell Biol. 143 (1998) 991–1002.
[88] J.D. Dodge, Phycologia 14 (1975) 253–263.
[89] T. Osafune, J.A. Schiff, E. Hase, Cell Struct. Funct. 15 (1990) 99–105.
[90] A. Rikin, S.D. Schwartzbach, Proc. Natl. Acad. Sci. USA 85 (1988) 5117–5121.
[91] T. Osafune, S. Sumida, J.A. Schiff, E. Hase, J. Electron Microsc. 40 (1991) 41–47.
[92] T. Osafune, J. Schiff, E. Hase, Exp. Cell Res. 193 (1991) 320–330.
[93] J.A. Schiff, S.D. Schwartzbach, T. Osafune, E. Hase, J. Photochem. Photobiol. B Biol. 111 (1991) 219–236.
[94] C. Sulli, S.D. Schwartzbach, J. Biol. Chem. 270 (1995) 13084–13090.
[95] C. Sulli, S.D. Schwartzbach, Plant Cell 8 (1996) 45–53.

[96] R. Kishore, S.D. Schwartzbach, Plant Sci. 85 (1992) 79–89.
[97] U. Muchhal, S.D. Schwartzbach, Plant Mol. Biol. 18 (1992) 287–299.
[98] R.L. Chan, M. Keller, J. Canaday, J.H. Weil, P. Imbault, EMBO J. 9 (1990) 333–338.
[99] R.G. Hiller, P.M. Wrench, F.P. Sharples, FEBS Lett. 363 (1995) 175–178.
[100] R. Rowan, S. Whitney, A. Fowler, D. Yellowlees, Plant Cell 8 (1996) 55–564.
[101] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Protein Eng. 10 (1997) 1–6.
[102] R. Kishore, U.S. Muchhal, S.D. Schwartzbach, Proc. Natl Acad. Sci. USA 90 (1993) 11845–11849.
[103] C. Sulli, Z.W. Fang, U. Muchal, S.D. Schwartzbach, J. Biol. Chem. 274 (1999) 457–463.
[104] O. Emanuelsson, H. Nielsen, G. Von Heijne, Protein Sci. 8 (1999) 978–984.
[105] L.S. Shashidhara, S.H. Lim, J.B. Shackleton, C. Robinso, A.G. Smith, J. Biol. Chem. 267 (1992) 12885–12891.
[106] B. Martoglio, M.W. Hofmann, J. Brunner, B. Dobberstein, Cell 81 (1995) 207–214.
[107] M. Lefort-Tran, M. Pouphile, G. Freyssinet, B. Pineau, J. Ultrastruct. Res. 73 (1980) 44–63.
[108] G.I. McFadden, J. Eukaryot. Microbiol. 46 (1999) 339–346.
[112] T. Tengs, O.J. Dahlberg, K. Shalchian-Tabrizi, D. Klaveness, K. Rudi, C.F. Delwiche, K.S. Jakobsen, Mol. Biol. Evol. 17 (2000) 718–729.
[113] E. Schnepf, M. Elbrächter, Grana 38 (1999) 81–97.
[114] P.R. Gilson, G.I. McFadden, Proc. Natl. Acad. Sci. USA 93 (1996) 7737–7742.
[115] A.L. Sharif, A.G. Smith, C. Abell, Eur. J. Biochem. 184 (1989) 353–359.
[116] Q. Lin, L. Ma, W. Burkhart, L.L. Spremulli, J. Biol. Chem. 269 (1994) 9436–9444.
[117] M. Plaumann, B. Pelzer-Reith, W.F. Martin, C. Schnarrenberger, Curr. Genet. 31 (1997) 430–438.
[118] K. Henze, A. Badr, M. Wettern, R. Cerff, W. Martin, Proc. Natl Acad. Sci. USA 92 (1995) 9122–9126.
[119] R. Vacula, J.M. Steiner, J. Krajcovic, L. Ebringer, W. Loffelhardt, DNA Res. 6 (1999) 45–49.