

Plastid ultrastructure defines the protein import pathway in dinoflagellates

Nasha Nassoury, Mario Cappadocia and David Morse*

Département de Sciences Biologiques, Université de Montréal, 4101 Sherbrooke est, Montréal, Québec, Canada H1X 2B2

*Author for correspondence (e-mail: david.morse@umontreal.ca)

Accepted 26 March 2003
Journal of Cell Science 116, 2867-2874 © 2003 The Company of Biologists Ltd
doi:10.1242/jcs.00517

Summary

Eukaryotic cells contain a variety of different compartments that are distinguished by their own particular function and characteristic set of proteins. Protein targeting mechanisms to organelles have an additional layer of complexity in algae, where plastids may be surrounded by three or four membranes instead of two as in higher plants. The mechanism of protein import into dinoflagellate plastids, however, has not been previously described despite the importance of plastid targeting in a group of algae responsible for roughly half the ocean's net primary production. Here, we show how nuclear-encoded proteins enter the triple membrane-bound plastids of the dinoflagellate *Gonyaulax*. These proteins all contain an N-terminal leader sequence with two distinct hydrophobic regions flanking a region rich in hydroxylated amino acids (S/T). We demonstrate that plastid proteins transit through the Golgi in vivo, that the first hydrophobic region in the

leader acts as a typical signal peptide in vitro, and that the S/T-rich region acts as a typical plastid transit sequence in transgenic plants. We also show that the second hydrophobic region acts as a stop transfer sequence so that plastid proteins in Golgi-derived vesicles are integral membrane proteins with a predominant cytoplasmic component. The dinoflagellate mechanism is thus different from that used by the phylogenetically related apicomplexans, and instead, is similar to that of the phylogenetically distant *Euglena*, whose plastids are also bound by three membranes. We conclude that the protein import mechanism is dictated by plastid ultrastructure rather than by the evolutionary history of the cell.

Key words: Dinoflagellate, Plastid, Nuclear-encoded protein, Protein import, Signal peptide

Introduction

Protein targeting towards the proper subcellular compartments is vital for the cell. For compartments surrounded by two membranes, nuclear-encoded proteins contain an N-terminal leader sequence that allows their targeting and post-translational translocation through both envelope membranes (Bruce, 2000; Strub et al., 2000; Vothknecht and Soll, 2000). In the case of higher plant chloroplasts, the acquisition of a chloroplast transit sequence allows the return to the plastid of proteins encoded by genes that were present in the original prokaryotic endosymbiont and were subsequently transferred to the eukaryotic host cell nucleus (Martin and Schnarrenberger, 1997; Martin et al., 1998). In some algae, however, plastids may be surrounded by either three or four membranes. These are thought to result from secondary endosymbioses (where a host cell acquires an eukaryotic endosymbiont already equipped with plastids), a hypothesis supported by differences in the molecular phylogeny of plastid and nuclear genes (Delwiche and Palmer, 1996; Morden et al., 1992). The transfer of genes with a chloroplast transit sequence from the nucleus of the primary endosymbiont to the nucleus of the secondary endosymbiont requires additional signals for correct targeting, which has led to the suggestion that the final composition of the leader sequence might recapitulate the evolutionary history of the plastid (McFadden, 1999). This is supported by analyses of leader sequences of plastid-directed proteins in the four membrane-bound plastids of haptophytes,

cryosphytes, apicomplexans and diatoms, all of which contain an N-terminal hydrophobic signal peptide (to target proteins to the ER) followed by a transit sequence which can direct proteins synthesized in vitro to purified plant plastids (van Dooren et al., 2001).

Several mechanisms are known for directing proteins to four membrane-bound plastids. In one mechanism, such as that used by haptophytes and diatoms, ribosomes are attached directly to the outer plastid membrane. After passage of the first (outermost) membrane, termed the chloroplast ER (Gibbs, 1981), proteins either cross the second membrane through large pores, or traverse the subsequent intermembrane space inside transport vesicles (van Dooren et al., 2001). In contrast to this mechanism, targeting to the four membrane-bound plastids of the apicomplexans does not involve ribosomes bound to the outer membrane, and indeed, many aspects of the mechanism are still unknown (van Dooren et al., 2001). It is known that plastid proteins enter the ER membrane system using an N-terminal signal peptide, since constructs lacking the signal peptide produce a cytoplasmic protein (Waller et al., 2000). It is also known that a transit peptide, exposed after cleavage of the signal peptide, is required for entry into the plastid, as proteins lacking this transit peptide are secreted from the cell (Waller et al., 2000). However, several mechanisms have been suggested for targeting proteins to the outermost plastid membrane from the ER (van Dooren et al., 2001). One proposal is that all proteins entering the ER

eventually pass through the outermost plastid compartment, and those destined to remain in the plastid bind to a receptor for the transit peptide. Alternatively, proteins may be specifically targeted to the outer plastid membrane directly from the Golgi. While the observation that plastid directed proteins lacking a transit peptide are secreted suggests passage through the Golgi, it must be stressed that this has not been directly shown.

Intiguously, nuclear-encoded proteins destined for the triple membrane-bound plastids of *Euglena* clearly transit through the Golgi (van Dooren et al., 2001). Details of the mechanism differ, however, as in *Euglena* the plastid-directed leader sequences contain two hydrophobic regions separated by an S/T-rich region reminiscent of a plastid transit sequence (Henze et al., 1995; Kishore et al., 1993). In vitro studies have demonstrated that the plastid directed proteins remain inserted in the ER membranes, unlike the apicomplexans where the plastid proteins are soluble inside the ER (Kishore et al., 1993; Osafune et al., 1990; Sulli et al., 1999; Sulli and Schwartzbach, 1996). The second hydrophobic region present in the *Euglena* leaders is absent in apicomplexan leaders (Foth et al., 2003).

In contrast to the above, there are no reports describing the transport mechanism used by nuclear-encoded plastid proteins in dinoflagellates. This is an unfortunate gap in our understanding of protein import pathways as dinoflagellates, along with diatoms, constitute the bulk of the phytoplankton, and the oceans are responsible for roughly half the primary production of the biosphere (Field et al., 1998). The dinoflagellates contain members with several types of plastids, of which those containing the carotenoid peridinin are the most prevalent. The next most abundant type of plastids contains fucoxanthin, and these are proposed to share a common ancestor with the peridinin-containing plastids (Yoon et al., 2002). Both types of plastids are surrounded by three membranes, and may thus employ a similar mechanism for protein import. On the one hand, this mechanism might be expected to be similar to that of apicomplexans because of their close phylogenetic relationship to the dinoflagellates (Sogin et al., 1996). On the other hand, protein import into dinoflagellate plastids might also be similar to that in *Euglena*, because the leader sequences of dinoflagellate plastid-directed proteins (Fagan et al., 1998; Le et al., 1997) are similar to those used by *Euglena*.

To distinguish between these two possibilities, we have characterized the import pathway into peridinin-containing plastids of the dinoflagellate *Gonyaulax polyedra*. We have studied import of two proteins for which we have previously raised antibodies (Nassoury et al., 2001), the carbon fixing enzyme Ribulose biphosphate carboxylase/oxygenase (Rubisco) and the soluble light-harvesting peridinin-chlorophyll *a*-protein (PCP). Dinoflagellate Rubisco is unusual, in that it is a form II enzyme and is formed only of large 55 kDa subunits (Morse et al., 1995). This enzyme is synthesized as a polyprotein, which must be processed to form the mature large subunits by cleavage of small linker peptides (Rowan et al., 1996). The mechanism and location of this processing event is unknown, although the final destination of the protein is the stroma, as determined by immunolocalization (Nassoury et al., 2001). The soluble PCP is not a polyprotein, and it is located in the thylakoids, also as determined by immunolocalization (Nassoury et al., 2001). The synthesis of

both proteins is under control of the circadian clock in vivo, and their expression times differ slightly. When the algae are grown under a 12:12 light:dark (LD) regime, Rubisco is synthesized from about midnight to midday (LDT 18 to 6), while PCP is synthesized from dawn to dusk (LDT 0 to 12) (Markovic et al., 1996). We found that nuclear-encoded proteins passed through the Golgi on route to the plastid, as has been shown for *Euglena* (Schiff et al., 1991). Furthermore, we show that the dinoflagellate plastid-directed protein leader sequences are functionally similar to those found in *Euglena*. We suggest that the second hydrophobic region present in the leader is a mechanistic requirement dictated by the triple membrane architecture of the plastids.

Materials and Methods

Luciferase constructs

The PCP leader sequence was amplified by PCR from a complete genomic clone using the oligonucleotides 5'-agccATGGGCC-GCTCTCGT-3' and 5'-aacatggtGTCGGCGAAGGCTGC-3'. The restriction site *NcoI* (lower case) was incorporated into primers to assist cloning of the product into pSP-luc+fusion vector (Promega Biotech) in-frame at the 5' end of the luciferase coding sequence. The construct containing the complete PCP leader sequence and luciferase (1-2 LUC) was then cloned into the *BamHI* and *HindIII* sites of the vector pBluescript for further manipulations. For in vivo transformation, the 30 nucleotide 5'UTR of the *S. chacoense* S₁₁-RNase gene (Saba-El-Leil et al., 1994) was inserted at the *HindIII* site of pBluescript located at the 5' end of the PCP leader sequence by two sequential PCR amplifications. The first amplification used oligonucleotides 5'-TGTTCAAAGTCAAATGGGCCGCTCTCGTAC-3' and T3 on the 1-2 LUC construct. For the second, this fragment was purified and re-amplified using oligonucleotides 5'-ataagcctGT-ACGATGAACTAAATTGTTCAAAGTCAAAGTCAA-3' and T3. The resulting DNA fragment was then digested by *HindIII* and *BamHI* and cloned back into pBluescript. A deletion construct lacking the first hydrophobic region of the leader sequence (2LUC) was generated by PCR amplification from the 1-2 LUC using 5'-ggatatacGCGCCGCGCTGCCTCCA-3', which corresponds to the 5' end sequence following the first signal peptidase site in the leader sequence of PCP and 5'-aggatatacTTTGCAGTTTGAACAT-3', which corresponds to the 3' end of the 5'UTR of the *S. chacoense* S₁₁-RNase gene. The restriction site *EcoRV* (lower case) was incorporated into primers to assist cloning as well as to provide an ATG start codon. The amplification product was digested with *EcoRV* and re-ligated, resulting in a plasmid missing the first hydrophobic core of the leader sequence. All PCR amplifications were done using proofreading DNA polymerase and all constructs were sequenced at each step. Constructs cloned separately into *KpnI* and *SacI* sites of the vector pBIN19 containing the 35 S double enhancer were introduced into the highly regenerable G4 line of the wild potato *Solanum chacoense* by *Agrobacterium*-mediated transformation as described (Matton et al., 1997).

In vitro translation and processing experiments

Constructs 1-2LUC and 2LUC were translated using a T7 TnT-coupled rabbit reticulocyte system (Promega Biotech) containing ³⁵S-methionine (ICN) in the presence or absence of 4 µl canine pancreatic microsomal membranes (Promega) as described by the supplier. Protease protection assays were performed by incubating the translation mix with 0.1 mg/ml trypsin in the presence or absence of 0.5% Triton X-100. Sedimentation of membrane-bound translation products was performed by centrifugation in an airfuge using an A-100/30 fixed angle rotor for 1 hour at 100,000 g. Translation products

were visualized by autoradiography after separation on 12% SDS polyacrylamide gels and electrophoretic transfer onto nitrocellulose membranes.

Immunocytochemistry and transformation

Gonyaulax cells were cultured and prepared for electron immunocytochemistry as described, using antibodies directed against a *Gonyaulax* Rubisco expressed from a cDNA in bacteria or PCP purified by column chromatography from the algae as described (Nassoury et al., 2001). A 20 nm gold-labeled goat anti-rabbit was used for detection (Ted Pella). A similar procedure was followed for immunolabeling of *S. chacoense* cells, except that a commercial anti-luciferase (Promega Biotech) was used as a primary antibody. A 20 nm gold-labeled rabbit anti-goat was used for detection (Ted Pella).

Sequence analysis

The sequences used for multiple alignments were obtained from GenBank using the accession numbers (top to bottom in Fig. 2) X94549, U93077, AF298221, AJ009670, AF028561, D14702, X89768, L21904, X15743, X66617 and S53593. Sequences for hydrophobicity plots (in Fig. 4) were AF087139, AF028561 and L21904. All analyses were performed using MacVector software, including ClustalW multiple alignments and neighbour-joining phylogenetic analyses. Bootstrap values are given for 10,000 trees.

Results

Nuclear-encoded plastid proteins transit through the Golgi

We first used immunocytochemistry to localize nuclear-encoded plastid proteins at times during the light-dark cycle when protein synthesis rates are high (Markovic et al., 1996). In one series of experiments, an antibody directed against the stromal protein Ribulose biphosphate carboxylase/oxygenase (Rubisco) (Nassoury et al., 2001) was used. Cells were first harvested at dawn, also termed LDT 0 (the start of the light phase in a 12:12 light: dark cycle), a time when the *in vivo* synthesis rate of Rubisco is high (Markovic et al., 1996). There is label not only over the chloroplast, as expected, but also over the Golgi (Fig. 1A,B). To confirm this result, we also examined the labeling patterns obtained with an antibody directed against the thylakoid protein peridinin-chlorophyll *a*-protein (PCP) (Nassoury et al., 2001). Again, sections were prepared from cells harvested at LDT 0, when the

in vivo synthesis rate of PCP is high (Markovic et al., 1996). The immunostaining shows that both the chloroplast and the Golgi are labeled (Fig. 1C,D). Thus, identical results are obtained with plastid proteins located in two different sub-organellar compartments. Furthermore, as Rubisco is encoded

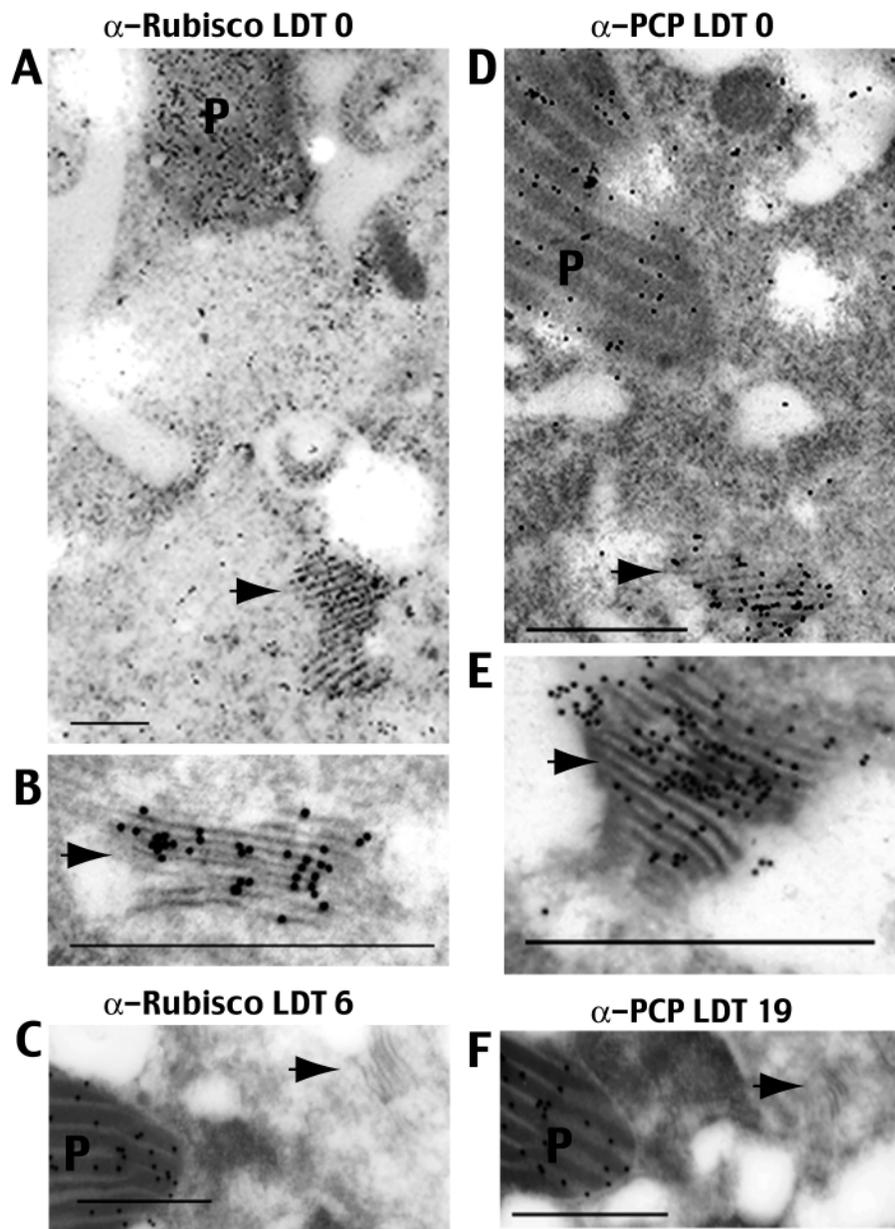


Fig. 1. Nuclear-encoded plastid proteins transit through the Golgi. *Gonyaulax* cell sections were treated with antibodies raised against two nuclear-encoded plastid proteins, Rubisco and peridinin-chlorophyll *a*-protein (PCP). In one series of experiments, cells were harvested and fixed at LDT 0 (by convention, the start of the light phase), a time when both proteins are actively synthesized *in vivo*. As a control for the specificity of the Golgi labeling, cells were also harvested and fixed at times when Rubisco or PCP were not actively synthesized *in vivo* (LDT 6 and 19, respectively). Sections were stained with either anti-Rubisco (A-C) or anti-PCP (D-F) as a primary antibody and a 20 nm gold-conjugated goat anti-rabbit as a secondary antibody. The Golgi is indicated by arrows in all pictures, and all scale bars represent 1 μ m. Note that Rubisco and PCP have different sub-organellar locations (stroma and thylakoid lumen, respectively) within the plastids (P). The thylakoid membranes appear white, as osmium tetroxide was not used during fixation to preserve the antigenicity of the proteins.

as a polyprotein (Rowan et al., 1996) while PCP is not (Le et al., 1997), different types of nuclear-encoded plastid protein transit through the Golgi in dinoflagellates. We conclude that all nuclear-encoded plastid proteins will arrive at the outer plastid membrane in the form of Golgi-derived vesicles as observed in *Euglena* (van Dooren et al., 2001).

The plastid targeting sequence contains a signal peptide. To rule out the possibility that the observed Golgi labeling was a non-specific response to the antibodies, we immunolabelled sections from cells isolated at midday (LDT 6) with the anti-Rubisco. At this time in their daily cycle, the *in vivo* synthesis rate of Rubisco is low (Markovic et al., 1996). We observed a constant labeling of the plastid (Fig. 1C), in agreement with previous results showing no difference in the total Rubisco levels detected on Western blots and by enzyme-linked immunosorbent assay at different times over the daily cycle (Nassoury et al., 2001). However, no label was found over the Golgi, confirming that the labeling seen at time of high synthesis rates was indeed due to the presence of Rubisco. This control experiment was also performed using cells isolated at midnight (LDT 19) and the anti-PCP antiserum. At this time, PCP synthesis rates are low (Markovic et al., 1996), and again, no labeling of the Golgi was observed (Fig. 1F). We conclude that both proteins must enter the Golgi from the ER at times of high synthesis rates.

Targeting to the ER typically requires the presence of an N-terminal hydrophobic signal peptide. Indeed, such a signal, followed by an AXA-type signal peptidase site, is found in the leader sequences of plastid directed proteins for both dinoflagellates and *Euglena*. In *Euglena*, the two hydrophobic regions have been proposed to act as start and stop signals for translocation across the ER membrane (Sulli et al., 1999). This topology results in a membrane-bound protein with the first signal peptidase site inside the ER lumen and the bulk of the protein outside.

We thus tested if dinoflagellate plastid leader sequences acted similarly to those studied in *Euglena*. For our *in vitro* analyses, we chose to examine the leader sequence of the thylakoid protein PCP. The leader sequence of this protein is virtually identical to that of the stromal protein glyceraldehyde-3-phosphate dehydrogenase (GAP) (Fagan et al., 1999), suggesting that the information to distinguish between stromal and thylakoid locations might be found within the coding sequence of the mature PCP. To avoid potentially conflicting signals, the complete PCP leader sequence, encoding both hydrophobic regions and the S/T-rich regions separating them (Le et al., 1997), was thus fused with a luciferase reporter gene. These constructs were transcribed and translated *in vitro* and the protein products analyzed by SDS PAGE. The size of the protein produced is as predicted when the transcripts are translated in the absence of canine microsomes, and is smaller when translated in the presence of canine microsomes. As expected, the processed form is larger than luciferase alone (Fig. 2A). We conclude that the size change upon addition of microsomes is consistent with cleavage of the signal peptide at the expected site.

This cleavage of the signal peptide following the signal peptide suggested that the first hydrophobic region is inserted into the microsomal membrane with the N-terminus in the

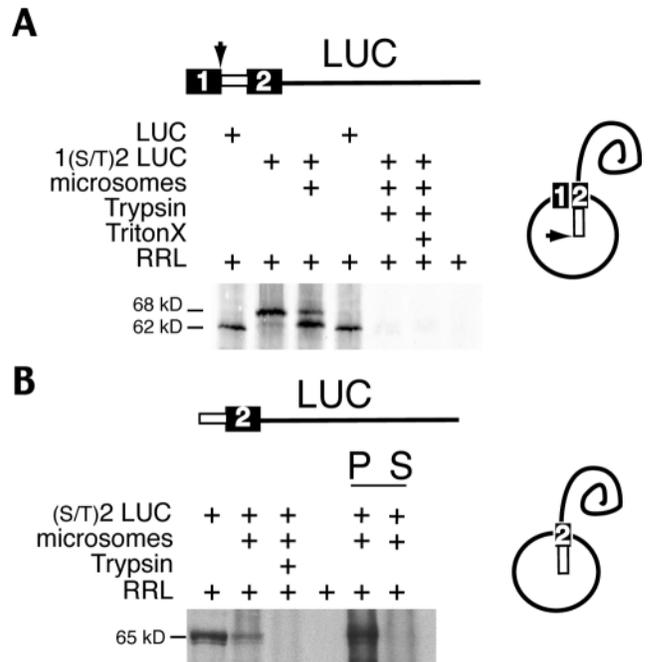


Fig. 2. The PCP leader contains a signal peptide that targets a reporter gene to canine microsomes *in vitro*. (A) A schematic view of the *G. polyedra* PCP leader used (top) shows two hydrophobic regions (numbered black boxes) separated by an S/T-rich region (white box) fused to luciferase (line). The small arrow indicates the potential AXA signal peptidase site. This construct was transcribed and translated *in vitro* in rabbit reticulocyte lysates (RRL) with the additions as illustrated. A smaller protein is produced after translation in the presence of canine microsomes, confirming cleavage after the first hydrophobic region. The topology of the protein, deduced from its susceptibility to trypsin digestion, is shown schematically on the right (small arrow indicates cleavage by the signal peptidase inside the vesicle). Note that the bulk of the protein remains outside the microsomal membrane. (B) A schematic view of the luciferase reporter fused with a modified *G. polyedra* PCP leader (top) lacking the first hydrophobic region. This construct was also translated as above. Most of the translation product is found in the pellet (P) rather than the supernatant (S) after ultracentrifugation, suggesting that it may be inserted into the microsomal membrane as found in *Euglena*. The predicted topology is again shown schematically on the right.

cytoplasm. If so, then the second hydrophobic region should act as a stop transfer signal, and the final topology of the protein, after cleavage by the signal peptidase, would be a single pass membrane protein with its C-terminal end in the cytoplasm. To test this, we digested with trypsin the protein produced in the presence of microsomes. This treatment reduced the level of radiolabelled protein, and no further decrease in the amount of protein could be obtained by inclusion of Triton X-100 in the digestion (Fig. 2A). This indicated that the small amount of signal resisting digestion did not represent luciferase inside the microsomes. The topology of the protein deduced from this experiment is unambiguous and is illustrated at right (Fig. 2A).

The leader contains a functional plastid transit sequence. These results also suggested that the second hydrophobic

region might act as a signal peptide if the first hydrophobic region was deleted, as has been observed in *Euglena* (Sulli et al., 1999). We thus prepared and translated a construct lacking the first hydrophobic region. The new construct showed no size decrease upon addition of microsomes, and the protein product was sensitive to trypsin digestion (Fig. 2B). To test the possibility that the hydrophobic region was inserted into the membrane, the labeled protein was centrifuged at 100,000 *g* for one hour following the translation. The labeled protein is found in the pellet, suggesting that the protein is either membrane bound or is tightly associated with the microsomal membrane. A likely topology for the protein, based on both these results and by analogy to results obtained with *Euglena*, is illustrated in the schema at right (Fig. 2B).

The topology of the plastid-directed proteins deduced from the *in vitro* experiments described above suggests that plastid precursor proteins may arrive at the plastid anchored in the membrane of a transport vesicle with their S/T-rich region inside the vesicle lumen. Fusion of the vesicle with the outer plastid membrane would make the S/T-rich sequence available to protein translocators spanning the inner two plastid membranes. We thus predicted that S/T-rich sequence would be equivalent to the transit sequence of higher plant plastid proteins. To test this, we asked if the translocators in the double membrane-bound plastids of higher plants could recognize this sequence *in vivo*. *In vivo* experiments allow a potential import into plastids to be evaluated in the context of competition with other compartments, unlike *in vitro* tests where only import into isolated pea chloroplasts is assessed (DeRocher et al., 2000; Wastl and Maier, 2000).

We prepared a reporter gene construct containing the leader sequence lacking the first hydrophobic region. This gene fusion, containing a *Solanum chacoense* 5'UTR and driven by the constitutive 35S promoter, was used to transform the wild potato *S. chacoense* (Matton et al., 1997). Leaves of eight transformed plants were screened and two selected for analysis on the basis of high signal with the antibody. The anti-luciferase antibody detects only the luciferase and has no reaction with untransformed plants and plants with low expression of the transgene (Fig. 3C). However, in the highly

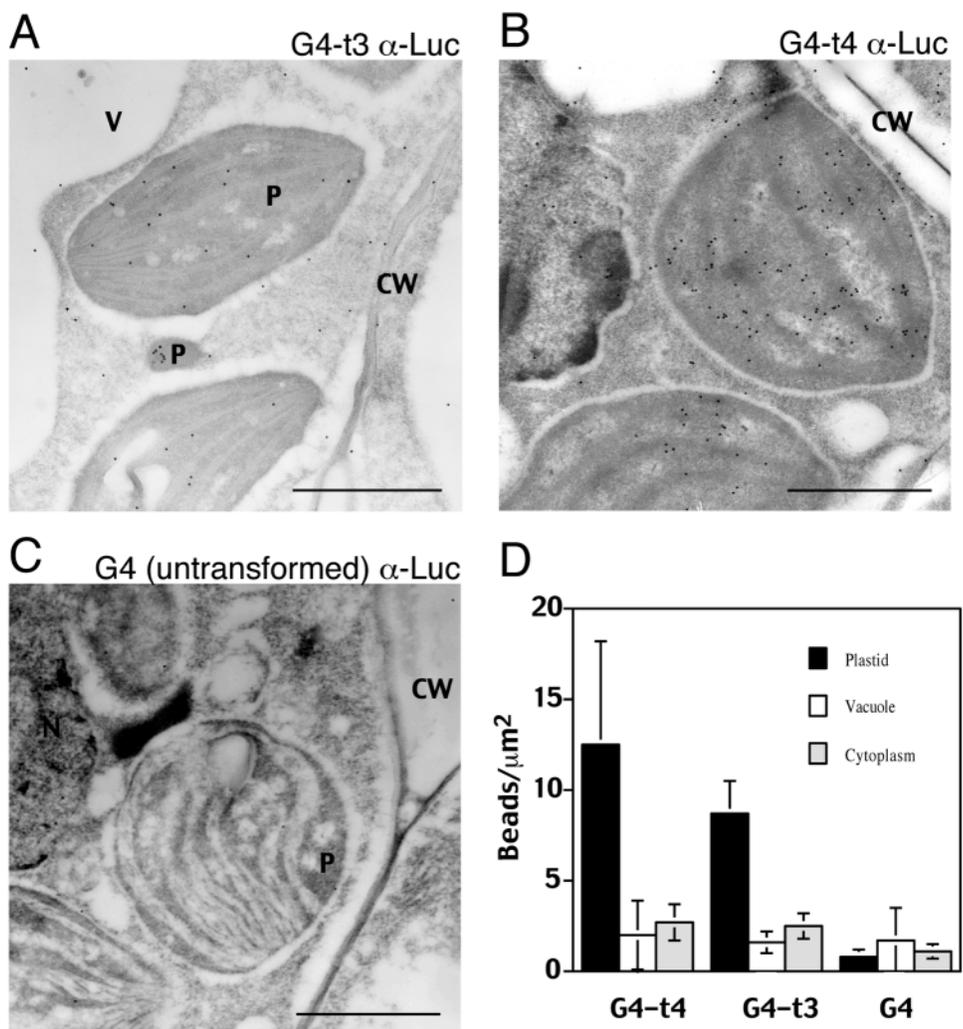


Fig. 3. Targeting to higher plant plastids by the S/T-rich region in the PCP leader. The luciferase reporter construct lacking the first hydrophobic region (as described in Fig. 2B) was fused with a *Solanum chacoense* 5'UTR and introduced into a highly regenerable genotype (G4) of *S. chacoense*. Two independent transformants expressed high levels of luciferase in the chloroplasts (A,B), as shown by immunoelectron microscopy using a commercial anti-luciferase as a primary antibody and a 20 nm gold-conjugated rabbit anti-goat as a secondary antibody. Label is observed over the plastid (P) but not the cell wall (CW), vacuole (V) or nucleus (N). Untransformed plants (C) show only background labeling. (D) Quantification of the label density is shown as number of gold beads per μm^2 for the two transformed plants above as well as for an untransformed control. Bars, 1 μm .

expressing plants, the subcellular location of the fusion protein was clearly the chloroplast. The presence of the reporter in the plastids confirms that the S/T-rich region acts as an authentic plastid targeting sequence in higher plants as well (Fig. 3A,B). To quantify the immunolabeling, five random pictures were taken from each sample, and the number of gold beads determined for plastid, vacuolar and cytoplasmic compartments. The label density in the plastids of transformed plants is 5- to 10-fold greater than that in other compartments within the same plant, and over 10-fold greater than the background labeling observed in the plastids of untransformed plants (Fig. 3D). Presumably, this conservation of function reflects the common evolutionary ancestor to all extant plastids. We have not observed staining in the ER or nuclear

membranes in these experiments. However, we cannot completely rule out some ER targeting, as we did not detect any signal above background in plants transformed with a luciferase reporter fused to the full-length leader sequence. If proteins directed to the ER membrane in this manner were unstable, for example, a direct comparison of the staining intensity in the two compartments would be impossible. In any event, it is clear that the S/T-rich region acts as an authentic transit peptide in higher plants.

Discussion

In this report we have addressed the protein import pathway used by dinoflagellates to allow entry of nuclear-encoded proteins into their triple membrane-bound plastids. We have shown that plastid-directed proteins transit through the Golgi in vivo (Fig. 1), and that plastid-directed proteins are directed to microsomes in vitro (Fig. 2). Protease sensitivity of these proteins indicates they have a topology similar to that found in *Euglena* (Sulli et al., 1999), where the bulk of the protein is cytoplasmic and the S/T-rich region of the leader is in the ER lumen. Lastly, we demonstrate that this S/T-rich region is a functional transit sequence, as it can direct the luciferase reporter into the plastids of higher plants (Fig. 3). This represents the first reported in vivo demonstration that an algal plastid transit sequence can function in higher plants. Taken together, these experiments characterize the key elements in dinoflagellate plastid protein targeting mechanisms, and thus fill a large gap in our understanding of protein entry into the plastids derived from secondary endosymbioses.

Similar to the leader sequences used to target other complex plastids, the dinoflagellate leader contains a hydrophobic N-terminal signal peptide and an internal plastid transit sequence. Unlike most other leader sequences, the dinoflagellate sequences contain a second hydrophobic region at the C-terminal end. This unusual feature is also found in *Euglena*, as shown by an alignment of five dinoflagellate sequences and six *Euglena* sequences (Fig. 4A). This alignment is color coded for visualization of the hydrophobic character (blue boxes). An AXA-type signal peptidase site follows the first hydrophobic regions (arrow) here manually aligned for comparison. The serine/threonine-rich regions (yellow boxes) that constitute the plastid transit sequence separate the two hydrophobic regions (Fig. 4A). Since *Euglena* and the dinoflagellates share a triple bound plastid architecture, we suggest that the C-terminal hydrophobic region, and the unusual protein topology that results from its insertion in the membrane of the transport vesicles, are required by the organelle's structure.

It is instructive to consider the dinoflagellate targeting mechanism in the light of the phylogenetic relationships between the different algal classes. First, it is generally accepted that a close phylogenetic relationship exists between apicomplexans and dinoflagellates (Fast et al., 2002; Sogin et al., 1996). This relationship can be illustrated by the molecular phylogeny of glyceraldehyde-3-phosphate dehydrogenase (GAP). Many examples of these trees are available in the literature, and the neighbor-joining tree shown here (Fig. 4B) is illustrative of much larger taxonomic samplings (Fast et al., 2001). These trees show the close relationship between apicomplexan and dinoflagellate hosts, as well as the relationship between euglenoids and trypanosomes. An

interesting feature of the GAP phylogeny is the clustering of plastid directed sequences in the apicomplexans and the dinoflagellates (Fig. 4B) (Fast et al., 2001). This phylogeny is supported by some plastid sequences (Fast et al., 2002; Wilson et al., 1996; Zhang et al., 2000), but not by a *tufA* phylogeny (Kohler et al., 1997) or the presence of a green algal-like mitochondrial *cox2* gene in the apicomplexan nucleus (Funes et al., 2002) that suggest green algal ancestry for the apicoplast. In any event, plastids from dinoflagellates and *Euglena* have had different evolutionary origins, even though both have three membranes and a second hydrophobic region in the leader sequences (Fig. 4B). In the model discussed below, we consider the case where dinoflagellate and apicomplexan plastids share a common ancestor. However, it is important to note that similar conclusions can also be reached in other scenarios, provided the plastids of dinoflagellates initially had four bounding membranes.

One possible evolutionary scheme, based on the GAP phylogeny shown (Fig. 4B), would involve uptake of a red algal symbiont by the common ancestor to both dinoflagellates and apicomplexans (Fig. 4C). Genes that had been transferred to the red algal nucleus, either before or after this secondary endosymbiotic event, would then be transferred to the new host cell nucleus. Given the ability of the dinoflagellate S/T-rich transit sequences to still function in higher plants (Fig. 3), any sequences transferred to the new host nucleus presumably included this plastid transit sequence. To re-enter the plastids, now bounded by four membranes, a signal peptide would have to be acquired in addition to the transit peptide (McFadden, 1999). Thus, according to this scheme, plastid-directed proteins in the last common ancestor for the apicomplexans and dinoflagellates would be expected to have both a signal peptide and a transit peptide in their leaders. The divergence of the dinoflagellates, and the loss of one bounding membrane around their plastids, must then have been accompanied by the selection of modifications in the C-terminal end of the leader sequence with an accentuated hydrophobicity. This new hydrophobic region, common to the sequence of plastid-directed nuclear proteins of dinoflagellates and *Euglena* (Fig. 4A), thus appears to be an essential element for targeting nuclear-encoded proteins to triple membrane-bound plastids.

The scheme above posits a transfer of genes to the nucleus of the new host cell before loss of the outermost plastid membrane. So far, however, it has not been possible to date the timing of these events. An alternative model, in which membrane loss occurred prior to gene transfer, would require hydrophobic signal peptide-like sequences to be inserted both before and after the transit peptide. A third model, where genes were transferred from the plastid genome directly to the new host nucleus, is not favored because of the fact that the reduced gene content of most extant plastids shows considerable overlap (Palmer and Delwiche, 1996). In any event, the peptide leader sequence does not recapitulate the gene transfer process, as at least part of the leader is defined not by phylogeny but by the mechanistic requirements of protein targeting to triple membrane-bound plastids.

What mechanistic requirement may be filled by the second membrane-spanning region in the dinoflagellate leader? One possible role may be to present the plastid transit sequence to the protein translocators in the inner (second) membranes. However, it may also act as a recovery mechanism should any

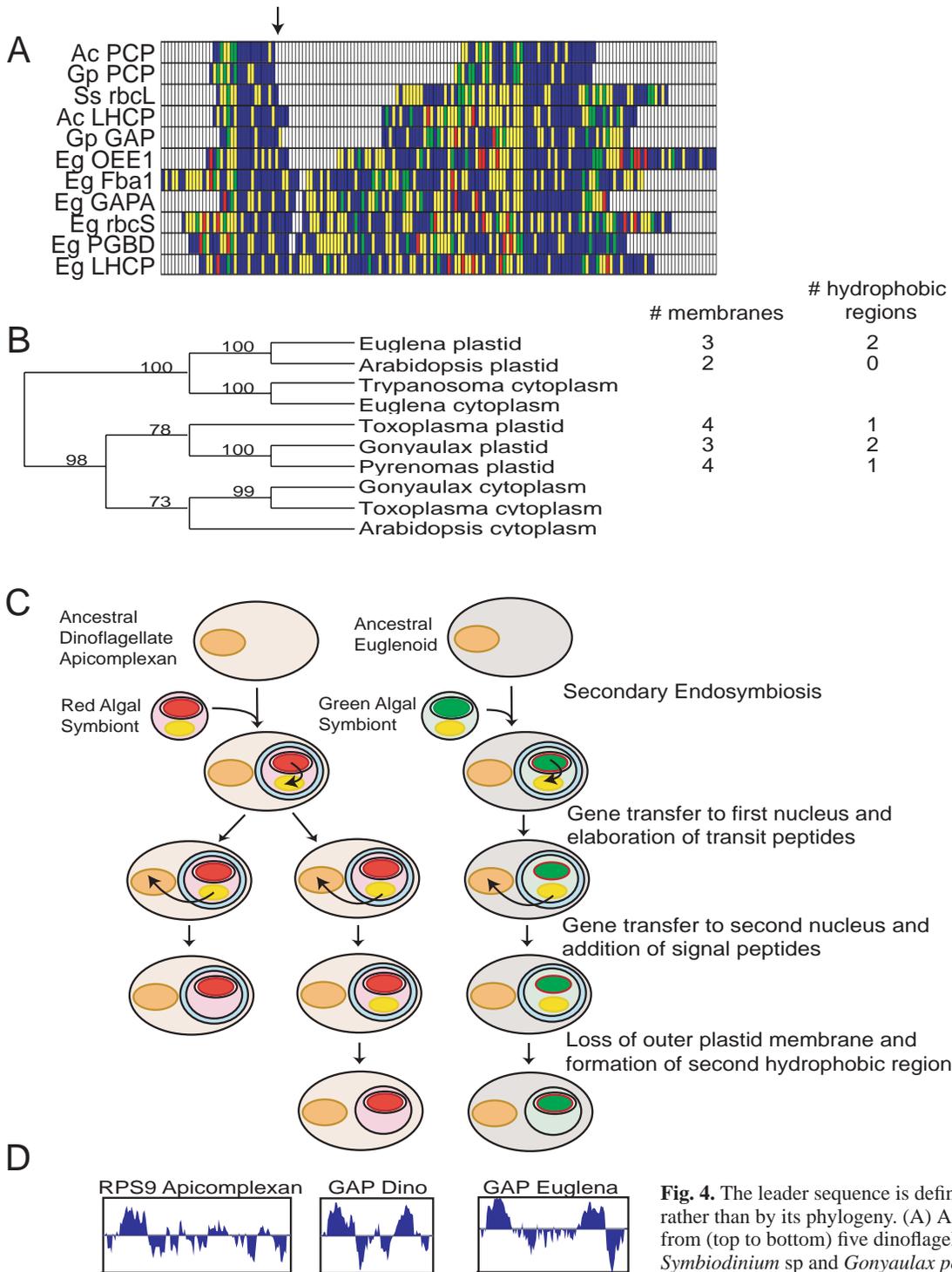


Fig. 4. The leader sequence is defined by plastid ultrastructure rather than by its phylogeny. (A) Alignment of signal peptides from (top to bottom) five dinoflagellate (*Amphidinium carterae*, *Symbiodinium* sp and *Gonyaulax polyedra*) and six *Euglena gracilis* plastid-directed, nuclear-encoded proteins. Two

conserved hydrophobic regions (blue) are interspersed with polar (yellow), acidic (red) or basic (green) residues. Gaps (white) were introduced to align the hydrophobic regions. The AXA signal peptidase site is at the end of the first hydrophobic region (arrow). (B) A phylogenetic reconstruction using glyceraldehyde-3-phosphate dehydrogenase sequences shows significant bootstrap support for the apicomplexan *Toxoplasma* and the dinoflagellate *Gonyaulax* as sister clades, using both plastid and cytoplasmic isoforms, as well as strong bootstrap support excluding *Euglena*. Numbers at the nodes indicate bootstrap support (10,000 trees). (C) Plastid evolution. Dinoflagellates and apicomplexans share a common ancestor both for the host cell and the plastid, and are unrelated to either Euglenoids or their plastids. Gene transfer from the endosymbiont nucleus to the new host cell nucleus must be accompanied by addition of a peptide signal, which allows the plastid-directed proteins to pass the new membranes. (D) Representative hydrophobicity plots of signal peptide sequences from the apicomplexan ribosomal protein S9 (RPS9), showing only a single hydrophobic region, as well as from a dinoflagellate and a Euglenoid glyceraldehyde-3-phosphate dehydrogenase (GAP), showing two distinct hydrophobic regions. All plots are to the same scale, with hydrophobicity increasing above the midline.

of the plastid proteins be misdirected to the plasma membrane, as soluble proteins secreted from the cell would be irretrievably lost. Indeed, as little is known about the targeting mechanism that delivers proteins specifically to the plastid, this now represents the next challenge in elucidating the protein import mechanisms in dinoflagellates and *Euglena*.

We thank L. Pelletier for technical assistance with the electron microscope, G. Laublin for expertise in plant transformation and tissue culture, K. K. Rashidan for technical assistance with the airfuge, S. Gibbs for helpful discussions and interpretation of the immunoelectron micrographs, and D. T. Luu for critical review of the manuscript. We also thank two anonymous reviewers for thoughtful comments and effective suggestions on figure organization. This work has been supported by the National Sciences and Engineering Research Council of Canada (to D.M. and M.C.).

References

- Bruce, B. D.** (2000). Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol.* **10**, 440-447.
- Delwiche, C. and Palmer, J.** (1996). Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plasmids. *Mol. Biol. Evol.* **13**, 873-882.
- DeRocher, A., Hagen, C. B., Froehlich, J. E., Feagin, J. E. and Parsons, M.** (2000). Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system. *J. Cell Sci.* **113**, 3969-3977.
- Fagan, T., Hastings, J. and Morse, D.** (1998). Phylogeny of Glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates. *J. Mol. Evol.* **47**, 633-639.
- Fagan, T., Morse, D. and Hastings, J.** (1999). Circadian synthesis of a nuclear encoded chloroplast Glyceraldehyde-3-phosphate dehydrogenase in the dinoflagellate *Gonyaulax polyedra* is translationally controlled. *Biochemistry* **38**, 7689-7695.
- Fast, N. M., Kissinger, J. C., Roos, D. S. and 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.
- Fast, N. M., Xue, L., Bingham, S. and Keeling, P. J.** (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. *J. Eukaryot. Microbiol.* **49**, 30-37.
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. and Falkowski, P.** (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237-240.
- Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S., Cowman, A. F. and McFadden, G. I.** (2003). Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science* **299**, 705-708.
- Funes, S., Davidson, E., Reyes-Prieto, A., Magallon, S., Herion, P., King, M. P. and Gonzalez-Halphen, D.** (2002). A green algal apicoplast ancestor. *Science* **298**, 2155.
- Gibbs, S.** (1981). The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann. N. Y. Acad. Sci.* **361**, 193-208.
- Henze, K., Badr, A., Wettern, M., Cerff, R. and Martin, W.** (1995). A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl. Acad. Sci. USA* **92**, 9122-9126.
- Kishore, R., Muchhal, U. S. and Schwartzbach, S. D.** (1993). The presequence of *Euglena* LHCPII, a cytoplasmically synthesized chloroplast protein, contains a functional endoplasmic reticulum-targeting domain. *Proc. Natl. Acad. Sci. USA* **90**, 11845-11849.
- Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D. and Roos, D. S.** (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science* **275**, 1485-1489.
- Le, Q. H., Markovic, P., Jovine, R., Hastings, J. and Morse, D.** (1997). Sequence and genomic organization of the peridinin-chlorophyll a-protein from *Gonyaulax polyedra*. *Mol. Gen. Genet.* **255**, 595-604.
- Markovic, P., Roenkeberg, T. and Morse, D.** (1996). Phased protein synthesis at several circadian times does not change protein levels in *Gonyaulax*. *J. Biol. Rhythms* **11**, 57-67.
- Martin, W. and Schnarrenberger, C.** (1997). The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32**, 1-18.
- Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M. and Kowallik, K. V.** (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Annu. Rev. Genet.* **393**, 162-165.
- Matton, D. P., Maes, O., Laublin, G., Xike, Q., Bertrand, C., Morse, D. and Cappadocia, M.** (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. *Plant Cell* **9**, 1757-1766.
- McFadden, G. I.** (1999). Plastids and protein targeting. *J. Eukaryot. Microbiol.* **46**, 339-346.
- Morden, C., Delwiche, C., Kuhsel, M. and Palmer, J.** (1992). Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems* **28**, 75-90.
- Morse, D., Salois, P., Markovic, P. and Hastings, J. W.** (1995). A nuclear encoded form II rubisco in dinoflagellates. *Science* **268**, 1622-1624.
- Nassoury, N., Fritz, L. and Morse, D.** (2001). Circadian changes in ribulose-1,5-bisphosphate carboxylase/oxygenase distribution inside individual chloroplasts can account for the rhythm in dinoflagellate carbon fixation. *Plant Cell* **13**, 923-934.
- Osafune, T., Yokota, A., Sumida, S. and Hase, E.** (1990). Immunogold localization of ribulose-1,5-bisphosphate carboxylase with reference to pyrenoid morphology in chloroplasts of synchronized *Euglena gracilis* cells. *Plant Physiol.* **92**, 802-808.
- Palmer, J. D. and Delwiche, C. F.** (1996). Second-hand chloroplasts and the case of the disappearing nucleus. *Proc. Natl. Acad. Sci. USA* **93**, 7432-7435.
- Rowan, R., Whitney, S. M., Fowler, A. and Yellowlees, D.** (1996). Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *Plant Cell* **8**, 539-553.
- Saba-El-Leil, M., Rivard, S., Morse, D. and Cappadocia, M.** (1994). The S11 and S13 self incompatibility alleles in *Solanum chacoense* Bitt. are remarkably similar. *Plant Mol. Biol.* **24**, 571-583.
- Schiff, J. A., Schwartzbach, S. D., Osafune, T. and Hase, E.** (1991). Photocontrol and processing of LHCPII apoprotein in *Euglena*: possible role of Golgi and other cytoplasmic sites. *J. Photochem. Photobiol. B* **11**, 219-236.
- Sogin, M. L., Morrison, H. G., Hinkle, G. and Silberman, J. D.** (1996). Ancestral relationships of the major eukaryotic lineages. *Microbiologia* **12**, 17-28.
- Strub, A., Lim, J. H., Pfanner, N. and Voos, W.** (2000). The mitochondrial protein import motor. *Biol. Chem.* **381**, 943-949.
- Sulli, C. and Schwartzbach, S. D.** (1996). A soluble protein is imported into *Euglena* chloroplasts as a membrane-bound precursor. *Plant Cell* **8**, 43-53.
- Sulli, C., Fang, Z., Muchhal, U. and 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.
- van Dooren, G. G., Schwartzbach, S. D., Osafune, T. and McFadden, G. I.** (2001). Translocation of proteins across the multiple membranes of complex plastids. *Biochim. Biophys. Acta* **1541**, 34-53.
- Vothknecht, U. C. and Soll, J.** (2000). Protein import: the hitchhikers guide into chloroplasts. *Biol. Chem.* **381**, 887-897.
- Waller, R. F., Reed, M. B., Cowman, A. F. and McFadden, G. I.** (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J.* **19**, 1794-1802.
- Wastl, J. and Maier, U. G.** (2000). Transport of proteins into cryptomonads complex plastids. *J. Biol. Chem.* **275**, 23194-23198.
- Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W. et al.** (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261**, 155-172.
- Yoon, H. S., Hackett, J. D. and Bhattacharya, D.** (2002). A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc. Natl. Acad. Sci. USA* **99**, 11724-11729.
- Zhang, Z., Green, B. R. and Cavalier-Smith, T.** (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. *J. Mol. Evol.* **51**, 26-40.