In vivo characterization of diatom multipartite plastid targeting signals

Kirk E. Apt[1,*], Lioudmila Zaslavkaia[1], J. Casey Lippmeier[1], Markus Lang[2], Oliver Kilian[2], Rick Wetherbee[3], Arthur R. Grossman[4] and Peter G. Kroth[2,*]

[1]Martek Biosciences Corp, 6480 Dobbin Rd., Columbia, MD 21045, USA
[2]Fachbereich Biologie, Universität Konstanz, Postfach M611, 78457 Konstanz, Germany
[3]School of Botany, University of Melbourne, Parkville, Victoria 3052, Australia
[4]Carnegie Institution of Washington, Department of Plant Biology, 260 Panama Street, Stanford, CA 94305, USA.
*Authors for correspondence (e-mail: kirkapt@martekbio.com; peter.kroth@uni-konstanz.de)

Accepted 23 July 2002
Journal of Cell Science 115, 4061-4069 © 2002 The Company of Biologists Ltd
doi:10.1242/jcs.00092

Summary

Plastids of diatoms and related algae are delineated by four membranes: the outermost membrane (CER) is continuous with the endoplasmic reticulum while the inner two membranes are homologous to plastid envelope membranes of vascular plants and green algae. Proteins are transported into these plastids by pre-sequences that have two recognizable domains. To characterize targeting of polypeptides within diatom cells, we generated constructs encoding green fluorecent protein (GFP) fused to leader sequences. A fusion of GFP to the pre-sequence of BiP [an endoplasmic reticulum (ER)-localized chaperone] resulted in accumulation of GFP within the ER; a construct encoding the pre-sequence of a plastid protein fused to GFP was directed into the plastids. Additional constructs demonstrated that the N-terminal region of the bipartite plastid targeting pre-sequence was necessary for transport of polypeptides to the lumen of the ER, while the C-terminal region was shown to enable the proteins to traverse the plastid double envelope membrane. Our data strongly support the hypothesis of a multi-step plastid targeting process in chromophytic algae and raises questions about the continuity of the ER and CER and the function of the latter in polypeptide trafficking.

Key words: Complex plastids, Bipartite pre-sequence, Diatom, Green fluorescent protein, Plastid import

Introduction

Diatoms (Bacillariophyceae) are a diverse and ubiquitous group of photoautotrophic, aquatic organisms. They are abundant primary producers and are responsible for fixing a significant portion of carbon in the oceans (Werner, 1977; Tréguer et al., 1995). Taxonomically, diatoms have been allied with the *Phaeophyceae* (brown algae) and *Chrysophyceae* within the kingdom *Chromista* (Cavalier-Smith, 1986; Cavalier-Smith, 2000) (familiarly called chromophytes). These algal groups are characterized by a photosynthetic apparatus that contains xanthophylls (fucoxanthin) and chlorophylls *a*/*c* as dominant light-harvesting pigments (Grossman et al., 1995). Diatoms and other chromophytic algae also possess 'complex plastids' delineated by four distinct membranes (Gibbs, 1979). This feature has been proposed to reflect the origin of chromophytic plastids by a secondary endocytobiotic event: that is, engulfment of a eukaryotic alga by a nonphotosynthetic eukaryote that led to permanent residence of the photosynthetic organism within the host cells and subsequent evolution of a plastid. The ancestral chromophytic plastid is thought to be related to red algae (Martin et al., 1998), which themselves arose by primary endocytobiosis. Organelle evolution appears to be linked to a massive transfer of genes from the genome of the engulfed photosynthetic organism to the host nuclear genome (Martin and Herrmann, 1998; Delwiche, 1999). While the endosymbiotic establishment of a plastid resulted in enhanced complexity of the chimeric cell, the evolution of a successful permanent residency required the modification of a number of cellular processes, such as the trafficking of both metabolites and proteins between the cytoplasmic and plastid compartments.

Four membranes delineate the chromophytic plastid. The inner two appear to be homologous to the double membrane envelope of plastids from red algae, green algae and vascular plants. A third unique membrane completely surrounds the plastid envelope and is thought to represent the former plasma membrane of the endosymbiont, while the outermost membrane may have evolved from the vacuolar/plasma membrane of the host organism. This outermost membrane has been observed to be continuous with the endoplasmic reticulum (Gibbs, 1981; Ishida et al., 2000). The portion of this ER-like membrane directly adjacent to plastids (and which appears to completely encase the plastid) is commonly referred to as the chloroplast ER or CER (Bouck, 1965).

Since many genes required for plastid function were lost from the genome of the endosymbiont and relocated to the nuclear genome of the host organism, it was critical to develop efficient transport of nuclear-encoded proteins across the four membranes that delineate the plastid. Early observations suggested that nuclear-encoded plastid polypeptides of diatoms and other chromophytic algae required a multi-signal pre-sequence to facilitate trafficking of polypeptides into plastids (Gibbs, 1979). Characterization of genes encoding several nuclear-encoded plastid polypeptides from chromophytic algae

(Grossman et al., 1990; Pancic and Strotmann, 1993) predicted pre-sequences with bipartite structures. The N-terminal regions of these pre-sequences are similar to classical signal sequences while the C-terminal regions have characteristics of transit peptides, sequences required for the transport of polypeptides into vascular plant plastids. In vitro studies demonstrated that the putative ER targeting domain could facilitate co-translational import of precursors of the fucoxanthin chlorophyll *a*/*c* binding proteins (FCP) into isolated canine microsomal vesicles (Bhaya and Grossman, 1991; Lang et al., 1998). Furthermore, the putative transit peptide region of the pre-sequence was sufficient to direct in vitro transport of polypeptides into isolated pea or spinach plastids (Lang et al., 1998). With the development of a relatively easy means of introducing genes into the nuclear genome of diatoms, we have started to dissect the role of specific domains of the pre-sequence associated with nuclear-encoded plastid polypeptides in the multi-step transport process that delivers these proteins to their site of function within the plastid.

Materials and Methods

Culture conditions

Phaeodactylum tricornutum Bohlin (University of Texas Culture Collection, strain 646) was grown at 20°C with continuous illumination at 75 μmol photon m^{-2} s^{-1} in Provasoli's enriched seawater medium (Starr and Zeikus, 1993) made with 0.5× Instant Ocean™ (I.O.) artificial seawater. Solid medium contained 1.2% agar and liquid cultures were bubbled with air containing 1% CO_2.

Plasmid constructs

All GFP fusions (Fig. 1) were inserted into the *Eco*RI and *Hind*III sites of the *P. tricornutum* transformation vector pPha-T1 (Zaslavskaia et al., 2000). This vector contains the *sh ble* (Zeo) gene fused with the *fcpB* promoter region and the *fcpA* terminator region. A multiple cloning site is flanked by the *fcpA* promoter region and the *fcpA* terminator region. Fusion of pre-sequences to EGFP were generated either by creating restriction sites by 'full-circle'-PCR or by fusion-PCR using fusion primers (36 bp) as described previously (Pont-Kingdon, 1997). For the C-terminal addition of the 6×His residues, individual PCR primers were designed containing a $(CAC)_6TAA$ motif added in frame 3′ of the last four codons of the GFP fusion genes. His-tagged polypeptides were partially purified from *P. tricornutum* transformants by fractionation of cells into soluble and membrane compartments. Soluble fusion protein was further purified by Ni-affinity chromatography (Qiagen NTA-column) and the isolated polypeptide sequenced from its N-terminus.

Microparticle bombardment

Cells were bombarded using the Bio-Rad Biolistic PDS-1000/He Particle Delivery System fitted with 1350 or 1500 psi rupture discs. The tungsten particles M17 (1.1 μm median diameter) were coated with 0.8 μg plasmid DNA in the presence of $CaCl_2$ and spermidine, as described by the manufacturer. Approximately 5×10^7 cells were spread in the center one third of a plate of solid 0.5× I.O. medium 1 hour before bombardment. The plate was positioned at the second level within the Biolistic chamber for bombardment. Bombarded cells were illuminated for 24 hours (cells divided once during this period) prior to suspension in 0.5 ml of sterile 0.5× I.O. medium; 100 μl of the cell suspension (~1×10^7 cells) were plated onto solid medium containing 100 μg/ml Zeocin. Plates were placed under constant illumination (75 μmol photon m^{-2} s^{-1}) for 2-3 weeks and resistant colonies re-streaked onto fresh solid medium containing 100 μg/ml Zeocin.

Fluorescence microscopy

Standard microscopical analyses and confocal laser scanning microscopy were performed using a Leica TCSNT system. Separation of GFP fluorescence and chlorophyll autofluorescence was achieved with a 525/550 nm band pass filter and a long pass filter of 590 nm. For 3D images, 50-90 individual scans of one cell were saved as a raytrace script build by the program 'KLM to POV', which served to reconstruct the 3D structure by the program 'POV-RAY'.

Electron microscopy

Cells of transformed *Phaeodactylum* strains were grown as clonal cultures in 250 ml erlenmeyer flasks containing 100 ml of sterile K-medium containing added silicates (Andersen et al., 1997). The cultures were maintained at 16°C in a 16 hour light/8 hour dark photoperiod under cool white and Grolux fluorescent lights. 50 ml aliquots of each strain were collected and centrifuged in 50 ml tubes at 1000 ***g*** for 5 minutes. The supernatant was discarded and droplets of the remaining cell suspension were sandwiched between two freezer hats (Type A, ProSciTech), with well depths of 100 μm, ensuring that all air bubbles were excluded. The enclosed cell suspensions were then frozen using the Leica EM High Pressure Freezer. The freezer hats enclosing the tissue samples were split apart, and the hats to which the frozen cell pellets were adhered were stored under liquid N_2 in cryovials prior to freeze-substitution. Frozen cell pellets were freeze-substituted in 0.1% uranyl acetate in acetone at –90°C for 72 hours, and the temperature raised to –50°C at 1°C/h. The cell pellets were scraped out of the freezer hats and rinsed three times for 30 minutes each in 100% acetone. Samples were then infiltrated with a graded series of HM20 low temperature resin in acetone consisting of 10% resin (5 hours), 30% resin (overnight), 50% resin (8 hours), 70% resin (12 hours), 90% resin (8 hours), and 100% resin (12 hours). The infiltrated samples were placed in a fresh change of 100% resin in gelatin capsules, polymerised under UV light for an additional 48 h at 50°C, and brought slowly to room temperature. The soft sample blocks were then hardened under UV light for a further 48 hours at room temperature. Polymerised blocks were sectioned on a Reichert Ultracut microtome and gold sections collected onto formvar coated 50 mesh hexagonal gold grids. Prior to immuno-gold-labelling, the sections on grids were blocked in PBS containing 0.8% BSA and 0.01% Tween-20 for 30 minutes. Grids were then incubated, section-side down, on 30 μl droplets of anti-GFP primary antibody, diluted 1:200 with blocking agent, for 3 hours at room temperature. The grids were washed four times on drops of blocking agent for 10 minutes each. Rinsed grids were then incubated on 30 μl droplets of goat anti-rabbit secondary antibody (diluted 1:20 with blocking agent), conjugated to 15 nm gold particles, for 12 hours at 4°C. Labelled grids were rinsed once on droplets of blocking agent, three times on droplets of PBS and then immersed twice for 30 seconds each time in distilled water. Negative controls were performed by using secondary antiserum only. The grids were then air dried and sequentially stained with uranyl acetate for 10 minutes and Triple Lead Stain for 5 minutes (Sato, 1968) and viewed in a Phillips Biotwin transmission electron microscope at 100 kV.

Results

Targeting sequences

To examine in vivo targeting of polypeptides to diatom ER or plastids, DNA sequences encoding the signal sequence of the ER lumenal chaperone BiP (Fig. 1A) or the pre-sequence of AtpC (γ subunit of plastid ATP synthetase) (Fig. 1B; AtpC1GFP) from the diatom *Phaeodactylum tricornutum*, as well as modified forms of the AtpC pre-sequence, were fused to the *GFP* gene (the constructs are designated AtpCGFP2-9, as shown in Fig. 1B). These gene fusions were inserted into

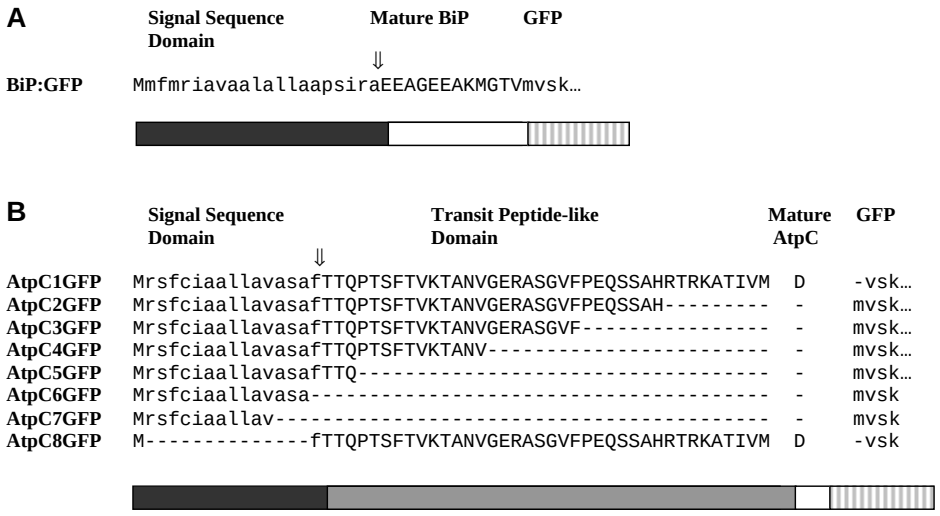

Fig. 1. Pre-sequences of fusion proteins of BiP (A) and AtpC (B) from *Phaeodactylum tricornutum* fused to GFP, and various fusions in which part of the AtpC pre-sequence was deleted. The signal sequences and the N-terminus of GFP are in small letters, while the transit peptide domains are shown in capital letters. The bars below represent the signal sequences (black), transit peptides (grey), N-termini of BiP and AtpC (white) and GFP (stripes). Arrows mark probable cleavage sites for signal sequences. Cleavage sites were determined by N-terminal sequencing, as described in the Results.

the multiple cloning site of the pPha-T1 vector (Apt et al., 1996; Zaslavskaia et al., 2000) and transformed into the diatom *P. tricornutum*.

The pre-sequence on the BiP:GFP chimeric protein used for these studies contained the complete signal sequence (21 amino acids) plus the first 12 amino acids of the mature protein (Fig. 1A). The pre-sequence of the nuclear-encoded AtpC protein from *P. tricornutum* is 54 amino acids (Fig. 1B) with a putative signal sequence from amino acids 1-16. The signal sequence domain is characterized by an arginine at position 2 followed by a region of hydrophobic amino acids. A transit peptide-like domain, with a high proportion of the hydroxylated amino acids serine and threonine, extends from amino acid 17 to 54. A similar transit peptide-like sequence is present on the unprocessed AtpC subunit of the centric diatom *Odontella sinensis* (Pancic and Strotmann, 1993) and on several other diatom plastid precursor proteins (P.G.K. and O.K., unpublished). All of the chimeric genes analyzed were expressed from the *fcpA* promoter, which was previously shown to effectively drive expression of GFP in *P. tricornutum* (Zaslavskaia et al., 2000; Zaslavskaia et al., 2001); localization of GFP in the cell was readily visualized by fluorescence microscopy.

ER targeting of GFP

Previous studies have demonstrated that GFP expressed in *P. tricornutum* becomes localized to the cytoplasm (Zaslavskaia et al., 2000; Zaslavskaia et al., 2001). As presented in the right panel of Fig. 2A, GFP devoid of a targeting sequence (vector designation pPTEGFP) appeared to accumulate in the cytoplasm, mostly towards the center of the cell in the region of the nucleus. The GFP fluorescence was excluded from plastids and vacuoles; the plastids are clearly visualized as brown-pigmented structures in the cell (Fig. 2A, left).

When a DNA fragment encoding the pre-sequence of the *P. tricornutum* ER-localized chaperone BiP (Apt et al., 1995) was fused to *GFP* and the chimeric gene (BiP:GFP) expressed in *P. tricornutum*, GFP fluorescence was observed in a network of membranes traversing the entire length of the cell (Fig. 2B, right). These membranes are thought to represent the ER. Addition of a diatom ER retention sequence [DDEL (Apt et al., 1995)] to the C-terminus of the BiP:GFP fusion protein made no observable difference in GFP localization (data not shown). Western blots did not show detectable amounts of GFP in the culture medium of BiP:GFP-expressing *Phaeodactylum* cells, indicating that also BiP:GFP without DDEL might not be secreted out of the cell. However, an effect of the DDEL domain on ER retardation of the BiP-GFP construct cannot be ruled out as GFP expression varies in independent transformants, and BiP-GFP without DDEL might be secreted/targeted to vacuoles followed by proteolytic degradation.

Confocal images of cells harboring BiP:GFP have a distinct sphere of GFP fluorescence near the middle of the cell (Fig. 3A) which, based on DAPI staining (not shown), corresponds exactly to the position of the nuclear envelope. The GFP fluorescence, in part, also delineates the position of the plastid. As shown in the 3D reconstitution of a confocal image series (Fig. 3B), a meshwork of GFP fluorescence extends throughout the cell (the position of the plastid was determined by red chlorophyll autofluorescence). There is the consistent appearance of ER membranes that associate with and form around the plastid, mostly longitudinal with respect to the plane of the cell. While results from electron microscopy suggest that diatom plastids reside within the lumen of the ER (see Introduction for references), the BiP signal sequence-targeted GFP was not observed in the compartment that isolates the plastid from the cytoplasm of the cell. These results suggest that detectable levels of BiP:GFP were not able to enter the CER segment of the intracellular membrane system.

To determine the exact site at which the BiP pre-sequence was cleaved, we constructed a vector encoding an N-terminal 6×His-tagged BiP:GFP fusion. This vector was transformed into *P. tricornutum* cells, and GFP was observed to localize to the ER in a pattern identical to that observed for the analogous construct devoid of the 6×His tag. The fusion protein was isolated from transformants by affinity chromatography using the Ni^{2+}-affinity column and the N-terminus of the purified polypeptide was sequenced. The results demonstrated that the transport of GFP into the ER involved cleavage of the chimeric polypeptide after alanine at position 21 (Fig. 1A).

Plastid targeting

Expression of GFP fused to the complete pre-sequence of the plastid-localized AtpC subunit (pATPC1GFP in Fig. 1B) resulted in accumulation of GFP in plastids; there is an exact congruence of GFP fluorescence with the position of the plastid as observed by both light (the plastid is apparent as a brown pigmented structure in Fig. 2C) and fluorescence microscopy (chlorophyll autofluorescence, not shown). The congruence was also visualized in the analysis of a confocal image series in which the relative position of GFP fluorescence was compared with that of chlorophyll autofluorescence (data not shown).

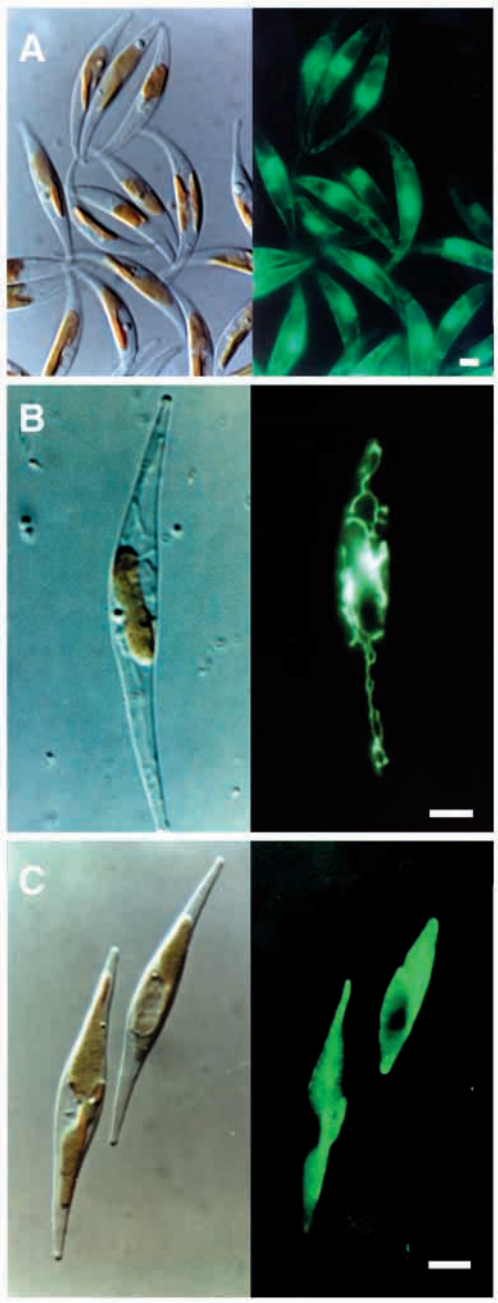

Fig. 2. *P. tricornutum* expressing GFP. Light microscopical images of cells are shown on the left while their corresponding fluorescence images (GFP) are presented on the right (excitation with UV light). (A) Cytosolic expression of GFP without a targeting sequence. (B) GFP targeted to the ER (BiPGFP construct). (C) GFP targeted to the plastids (AtpC1GFP construct). Bars, 2 μm.

Removal of the ER targeting domain of AtpC

In the construct AtpC8GFP, the region encoding the 14 amino acids following the initiator methionine at the N-terminus of the AtpC1GFP fusion protein was removed (Fig. 1B). GFP expressed from this construct accumulated in the cytoplasm of the cell. The pattern of accumulation was identical to that observed for GFP without a pre-sequence (Fig. 2A); no GFP fluorescence was apparent in either the plastid or ER. These results suggest that plastid-targeted polypeptides must enter the ER prior to plastid localization; the transit peptide-like domain alone is not sufficient for routing polypeptides into plastids in vivo.

Deletions of the transit peptide

While the size of the signal sequence is very similar among a variety of plastid pre-proteins characterized in chromophytes, the length of transit peptide-like domains may vary considerably. As in vascular plants, it is difficult to identify sequence motifs within transit peptides that might be critical for import function; however, all transit peptides (in vascular plants as well as in algae) appear to contain a high proportion of hydroxylated amino acids (von Heijne et al., 1989; Liaud et al., 1997). The AtpC1GFP chimeric gene was modified to generate deletions in the transit peptide segment of the pre-sequence. Constructs AtpC2GFP, AtpC3GFP and AtpC4GFP, shown in Fig. 1B, encode a series of AtpC:GFP deletions from the C-terminal end of the transit peptide domains. As many as 24 amino acids of the transit peptide could be removed, with 14 amino acids of the original transit peptide remaining, without an observable effect on plastid localization. Hence, the majority of the transit peptide-like domain of AtpC is not absolutely necessary for importing polypeptides into plastids. The addition of the putative ER retention sequence, DDEL, to the C-terminal ends of these chimeric polypeptides made no observable difference in the level of GFP that accumulated in plastids (data not shown). Surprisingly, the chimeric AtpC5GFP polypeptide, which contains only three amino acids of the presumed transit peptide (Fig. 1B), also accumulated in the plastid, although the efficiency of transport and accumulation is not known (the signal seems generally lower than in cells harboring the full length AtpC:GFP construct). Interestingly, when an ER retention signal was fused to the C-terminus of AtpC5GFP (AtpC5GFPDDEL construct), GFP fluorescence was visualized as strands of fluorescence extending over the length of the cells, and encircling the nucleus and plastids. This localization pattern appeared identical to that observed for cells expressing the BiP:GFP fusion (Fig. 2B, Fig. 3A), suggesting that fusion of the ER retention sequence to AtpC5GFP strongly interfered with the transfer of the precursor polypeptide from the ER to the plastid. Immunocytochemical localization of GFP in cells transformed with the AtpC5GFPDDEL construct confirmed that GFP was concentrated around the nuclear envelope (Fig. 4A) and the ER emanating from the envelope. In a number of sections, high levels of GFP were also visualized close to the plastid surface

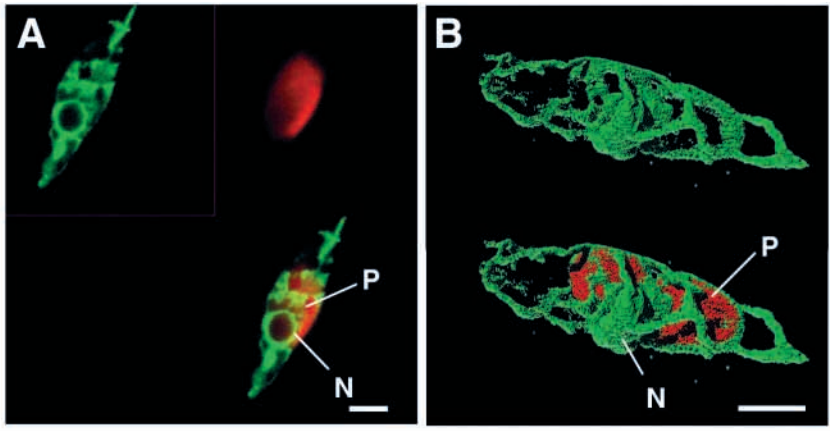

Fig. 3. GFP accumulation in the ER of *P. tricornutum* expressing a BiP:GFP fusion protein. (A) Images from a confocal laser scanning microscope. Top-left: green GFP fluorescence; Top-right: red chlorophyll fluorescence indicating the location of the plastid; bottom-right: overlay of the fluorescent images. (B) Reconstruction of 3D images from a series of confocal fluorescence micrographs of GFP fluorescence (top), and combined GFP and chlorophyll fluorescence (bottom). GFP is primarily located in cytosolic ER strands and the nuclear envelope. P, plastid; N, nucleus. Bars, 2 μm.

(Fig. 4B); localization at the plastid surface might reflect the position of the ER rings that were observed to encircle plastids in cells expressing BiP:GFP.

To determine the position at which the AtpC precursor protein is cleaved by the signal peptidase in vivo, an AtpC4GFP construct was 6×His tagged (fused to the C-terminus) and transformed into *P. tricornutum*. As in the case of the original AtpC4GFP construct, GFP expressed from the 6×His-tag-modified construct localized to the plastid (data not shown). The tagged polypeptide was purified from transformants by Ni^{2+} affinity chromatography and its N-terminal sequence was found to begin with TTQ. This placed the ER-recognized cleavage site between amino acid 16 (phenylalanine) and 17 (threonine) of the AtpC pre-sequence (Fig. 1B) and shows that cleavage by the stromal peptidase does not occur, probably because the recognition site for the protease was deleted in this construct.

Interestingly, in cells harboring AtpC4GFP fused to a 6×His tag, GFP fluorescence accumulated within a distinct region of the plastid near its mid-point (Fig. 5A). Immunocyto-chemistry of the transformed cells confirmed that the 6×His-tagged GFP was present in the plastid, and highly concentrated in a region corresponding to the 'lobe' that is positioned central to the plastid (Fig. 5B); the lobe region is associated with plastid division (Borowitzka et al., 1977; Borowitzka and Volcani, 1978). The presence of a 6xHis tag on plastid-targeted GFP may cause a spurious association of the polypeptide with components of the plastid division apparatus.

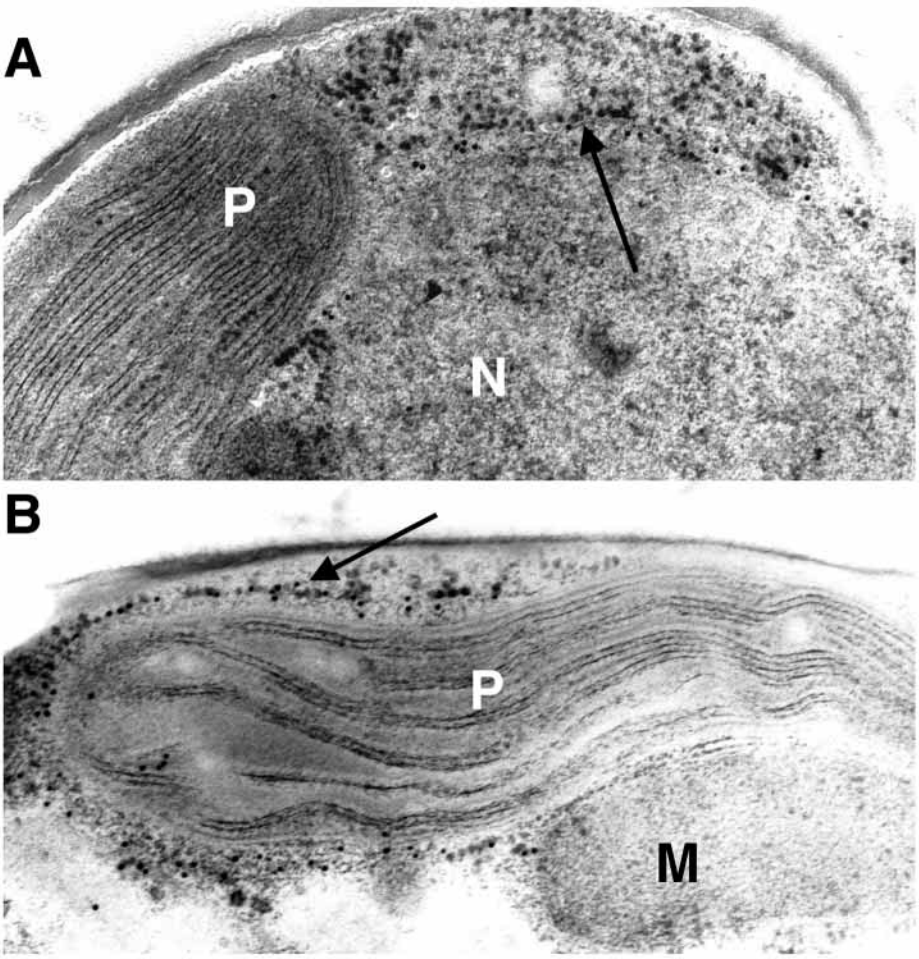

Fig. 4. Accumulation of GFP in the nuclear membrane (A) and on the plastid (B) surface of *P. tricornutum* expressing AtpC5-DDEL. Localization was demonstrated by immuno-electron microscopy using an antiserum against GFP (for details, see Materials and Methods). Arrows indicate areas that exhibit high GFP accumulation. P, plastid; N, nucleus; M, mitochondrion.

Deletions in the pre-sequence junction

Constructs AtpC6GFP and AtpC7GFP encode fusion proteins with deletions that extend from the C-terminal end of the transit peptide into the C-terminal end of the signal sequence (Fig. 1A). In both cases GFP fluorescence accumulated in structures near the center of the cell and adjacent to the plastids (Fig. 6A). Reconstitution of confocal images of strains transformed with AtpC6GFP and AtpC7GFP

showed that the centrally located GFP was spatially distinct from plastids and appeared to be within structures wrapped around the central portion of the plastid (Fig. 6B). Analysis of GFP localization by immunocytochemistry demonstrated that the GFP protein was found both in mitochondria and plastids, additional analysis using the fluorescent dye 'Mito-Tracker' supports this result (data not shown). In *P. tricornutum* a single, multi-lobed mitochondrium is typically present in the center of the cell and directly adjacent to the plastid. The extensive truncations of the AtpC pre-sequences associated with the AtpC6GFP and AtpC7GFP constructs apparently resulted in a highly anomalous localization pattern, with most GFP accumulating in mitochondria. Previous studies with yeast have demonstrated that a variety of randomly generated pre-sequences can promote the promiscuous entry of polypeptides into mitochondria (Allison and Schatz, 1986; Baker and Schatz, 1987).

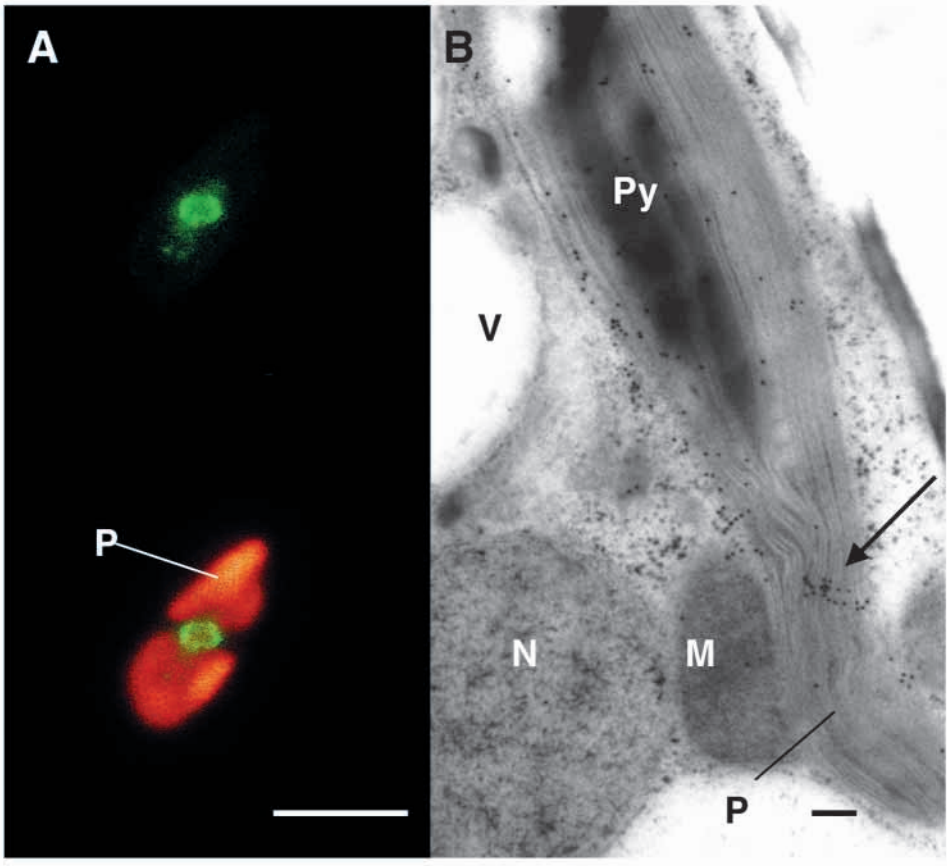

Fig. 5. Accumulation of His-tagged GFP (AtpC4-His) in the central region of the plastid in *P. tricornutum*. (A) Top, GFP fluorescence; bottom, superimposed GFP and chlorophyll fluorescence (the red chlorophyll fluorescence marks the two lobes of the plastids, P). The fusion protein appears to accumulate between the two plastid lobes. (B) Electron micrograph of a *P. tricornutum* cell immuno-decorated with antisera against GFP. GFP accumulates in the central region of the plastid (arrow), where two plastid lobes are connected. Py, pyrenoid; M, mitochondrium; N, nucleus; V, vacuole. Bars, 5 μm (A); 0.2 μm (B).

Discussion

Primary and secondary endocytobiosis occurred as a consequence of uptake by and their subsequent permanent establishment of prokaryotic and eukaryotic photoautotrophs, respectively, within heterotrophic hosts. While substantial evidence suggests that all plastids can be traced back to a single primary endocytobiosis (which led to the evolution of plastids in green algae, land plants, red algae and glaucophytes) (Delwiche and Palmer, 1997; Martin et al., 1998; Moreira et al., 2000), secondary endocytobiosis probably occurred several times as independent events, resulting in different algal lineages (chromophytic algae, cryptophytes, euglenophytes and dinoflagellates) (Delwiche, 1999) and some nonphotosynthetic organisms such as the *Plasmodium* or *Toxoplasma* parasites. These latter organisms contain organelles designated apicoplasts that represent reduced secondary plastids (Waller et al., 2000).

The evolution of organelles from free-living organisms following engulfment must have involved extensive

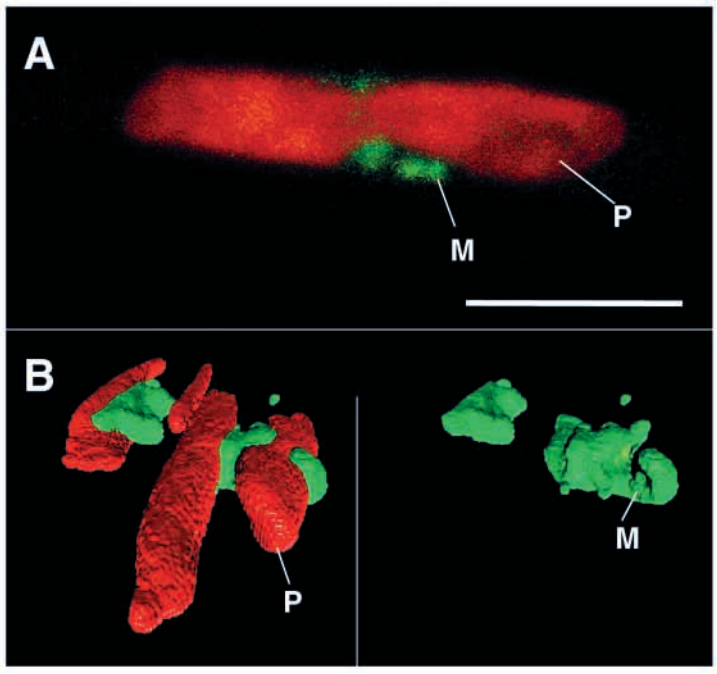

Fig. 6. Accumulation of GFP (AtpC6GFP) in mitochondria of *P. tricornutum*. (A) Confocal image of combined green GFP and red chlorophyll fluorescence. (B) 3D reconstruction of GFP and chlorophyll fluorescence of two dividing cells having two separated plastids each. Note the close association of plastids (P) and mitochondria (M). Bar, 7 μm.

transfer of genetic material from the genome of the endosymbiont to the nucleus of the host organism (Martin and Herrmann, 1998). Hence, the synthesis of many polypeptides required for diverse functions of the symbiont moved to the cytoplasm of the host cell. These events made it imperative that the associated organisms develop efficient pathways for importing polypeptides into evolving plastids. Recent information has added considerably to our understanding of the transport of polypeptides into plastids that are delineated by a double envelope membrane (evolved via primary endocytobiosis) (Chen and Schnell, 1999; Heins et al., 1998). In vascular plants and green algae, nuclear-encoded plastid polypeptides are translated in the cytosol of the cell and an N-terminal transit peptide directs post-translational import of polypeptides into plastids. In contrast, little is known about processes required for routing polypeptides into the complex plastids that have evolved via secondary endocytobiosis (van Dooren et al., 2001; Kroth, 2002).

Complex plastids have been observed in a variety of algal lineages. Plastids from these algae are delineated by multiple membranes (either three or four distinct membranes, with a CER only present in chromophytes and cryptophytes). Although strategies for routing polypeptides into complex plastids evolved following distinct endosymbiotic events, there are functional parallels among the different algal lineages with respect to mechanisms by which polypeptides are routed to plastids. All known pre-sequences necessary for the import of polypeptides into complex plastids are multipartite. For some organisms, the secretory system is involved in the primary targeting step (McFadden, 1999). In the *Euglenophytes*, the plastids are delineated by three membranes but are not associated with a CER (Gibbs, 1978). Cell fractionation studies have demonstrated an association of plastid pre-proteins with both Golgi and ER membranes, suggesting that precursor polypeptides initially pass through the cytosolic ER with subsequent translocation to the plastids via the Golgi system (Sulli et al., 1999). Secretory-system-dependent transport has also been demonstrated for *Plasmodium* (four membranes with no CER) (Waller et al., 2000).

Diatoms, a diverse group of organisms with four membranes that delineate the plastid (Gibbs, 1981), are representative of the chromophytic algae. The chromophytes include the *Bacillariophyceae* (diatoms), *Eustigmatophyceae*, *Phaeophyceae* and *Chrysophyceae*. Based on electron microscopy, plastids of chromophytic algae were proposed to be completely compartmentalized within the lumen of the ER, with those ER-like membranes immediately around the plastid being designated CER (Bouck, 1965; Gibbs, 1981). Gibbs suggested that polypeptides routed into chromophytic algal plastids must pass through the CER prior to translocation across the plastid envelope (Gibbs, 1981). This implied that targeting of polypeptides to plastids would require multiple targeting signals. In accord with this hypothesis, characterization of nuclear-encoded plastid polypeptides from various chromophytic algae were shown to have leader sequences with two putative targeting domains (Grossman et al., 1990; Apt et al., 1994; Apt et al., 1995; Kroth and Strotmann, 1999). The first domain resembles a classic signal sequence that appears to be required for transport into the ER (Walter and Johnson, 1994). The second domain of the pre-sequence is variable in length but has characteristics similar to those of transit peptides (Heins et al., 1998), which are enriched for the hydroxylated amino acids serine and threonine.

Some experiments have been performed to help establish the functions of the different pre-sequence domains associated with plastid-localized diatom polypeptides. These pre-sequences could drive co-translational import of proteins in vitro into canine microsomes (Bhaya and Grossman, 1991; Lang et al., 1998), a system extensively used to study the role of the signal sequence in transporting polypeptides into the ER. This work demonstrated that nuclear-encoded plastid polypeptides in the diatoms contain a functional ER targeting signal. Furthermore, the transit peptide-like domain, found on nuclear-encoded plastid polypeptides of diatoms, facilitated the import of polypeptides into isolated vascular plant plastids (Lang et al., 1998). Together, these studies illustrate that the different targeting signals present on nuclear-encoded plastid polypeptides of diatoms can facilitate the transport of those polypeptides across both ER and plastid envelope membranes. These results also demonstrate that no additional import signals within or at the C-terminus of the mature protein are necessary for the targeting process.

The generation of a complete in vitro system for elucidating events in the transport and processing of plastid polypeptides in the chromophytes will be extremely difficult as such a system would require cell disruption followed by the isolation of, or enrichment for plastids enclosed by four intact membranes. Wastl and Maier demonstrated that a preprotein encoded on the genome of the periplastidic nucleomorph of *Guillardia theta*, and therefore containing a transit peptide only, could be imported post-translationally into isolated homologous complex plastids (Wastl and Maier, 2000). Therefore, the outer membranes must have been lost or severely disrupted during the preparation of plastids.

To bypass the difficulties of developing an in vitro system, we have begun to examine targeting of polypeptides to diatom plastids, as well as to other subcellular compartments, using an in vivo system. In this in vivo system, constructs were generated that encode fusions between the pre-sequences of polypeptides targeted to different locations in the cell with GFP. In initial experiments we demonstrated that the complete pre-sequence of the AtpC precursor and of other plastid preproteins (O.K. and P.G.K., unpublished) are sufficient to drive the import of GFP into diatom plastids in vivo. Therefore, no other potential targeting signals within the mature region of diatom plastid precursors are required for transport of proteins across the four membranes that separate the stromal compartment of the plastid from the cytoplasm. The signal sequence-like domain of the AtpC pre-sequence plus three amino acids of the transit peptide fused to GFP directed GFP into the plastids. Removal of the signal peptide domain from the fusion protein resulted in cytoplasmic accumulation of GFP, while removal of most of the transit peptide, coupled with the addition of an ER retention signal (AtpC5GFPDDEL, Fig. 1), resulted in accumulation of GFP in the ER. The short transit peptide present on the AtpC5GFPDDEL construct did not prevent localization directed by the ER retention signal. These results strongly suggest that the first step in targeting polypeptides to plastids in chromophytic algae requires passage of that polypeptide into the ER. The GFP that accumulated in the ER appeared as thread-like strands present throughout the cytoplasm and extending laterally within the

cell. A sphere of fluorescence also encased the nucleus, probably representing the nuclear envelope. Immuno-electron microscopy confirmed that the GFP was localized to cytoplasmic ER and the nuclear envelope, which is continuous with cytoplasmic ER. Furthermore, a distinct ring of GFP fluorescence was shown to encircle plastids in the region of the valvar axis. A 3D reconstruction of confocal images demonstrates that this GFP ring does not completely encase the plastid. Based on EM studies of *P. tricornutum*, other diatoms and other chromophytes, an ER-like membrane called the CER is closely appressed to the plastid envelope membranes and completely surrounds the organelle (Crawford, 1973; Borowitzka and Volcani, 1978; Gibbs, 1981). The CER also appears to be continuous with cytoplasmic ER membranes. Furthermore, the periplastidic space between the CER and the plastid envelope membranes contains vesicular and tubular structures that have been hypothesized to be directly involved in transporting polypeptides from the ER to the outer envelope membranes of the plastid (Smith-Johannson and Gibbs, 1972; Gibbs, 1979; Gibbs, 1981). We never observed either ER- or plastid-targeted GFP in these vesicles (although it is possible that the small size of these vesicles and relatively low GFP content precluded their visualization). Hence the role of these vesicles in polypeptide trafficking remains uncertain.

The transit peptide domains in chromophytic algae are variable in both size and amino acid sequence. Diatom FCP pre-sequences have transit peptides of about 15 amino acids, while those of AtpC are 37 or 40 amino acids. Interestingly, our results indicate that a significant portion of the C-terminal region of the AtpC transit peptide domain is not required for targeting GFP to plastids (Table 1); as few as three amino acids (TTQ) at the N-terminus of this domain targeted GFP into plastids. Those regions that were deleted from the transit peptide might function to improve the efficiency of the import process.

Localization of GFP bearing AtpC pre-sequences appears to be unaffected by the presence of an ER retention signal, except in the case of the AtpC5GFP. GFP accumulated in plastids when transformants expressing the AtpC5GFP chimeric protein were devoid of an ER retention signal. However, when the DDEL sequence was attached to the C-terminus of this fusion protein, GFP accumulated in the ER. Since the DDEL sequence had no apparent effect on fusion proteins targeted to the ER by the BiP signal sequence or to plastids by the complete AtpC pre-sequence, it is unlikely that an ER retrieval mechanism accounts for the accumulation of the AtpC5GFPDDEL polypeptide in the ER. An altered conformation of this polypeptide as a consequence of the attached DDEL sequence may interfere with plastid routing. There may be no distinction between plastid- and ER-targeted polypeptides at the first import step. The exchange of polypeptides between the cytosol and plastid compartments may involve the entire network of ER membranes rather than just the CER. Routing of plastid-localized polypeptides through the general secretory pathway offers a potentially larger surface area for the exchange of substrate between plastids and cytosol, which may be advantageous for systems in which four membrane barriers constrain polypeptide trafficking. Other signals on plastid pre-proteins may function in the subsequent targeting of polypeptides from the CER, although there is no direct evidence that links the CER to the targeting process.

Table 1. Location of GFP accumulation as analyzed by light and confocal fluorescence microscopy

Construct	Without DDEL	With DDEL*
GFP	Cytoplasm	n.d.
BiP:GFP	ER	ER
BiP:GFPHis	ER	n.d.
AtpC1GFP	Plastid	n.d.
AtpC2GFP	Plastid	Plastid
AtpC3GFP	Plastid	Plastid
AtpC4GFP	Plastid	Plastid
AtpC4GFPHis	Plastid	n.d.
AtpC5GFP	Plastid	ER
AtpC6GFP	Mitochondria/plastid	n.d.
AtpC7GFP	Mitochondria/plastid	n.d.
AtpC8GFP	Cytoplasm	n.d.

The structures of individual constructs are shown in Fig. 1.
*With DDEL sequence added to the C-terminus of GFP (n.d., not determined).

The work presented in this manuscript demonstrates the power of using in vivo analyses to dissect polypeptide trafficking in diatoms. Along with other studies on protein targeting into complex plastids (Sulli et al., 1999; Waller et al., 2000), our results demonstrate the fundamental importance of the secretory pathway for the establishment and maintenance of secondary endocytobiosis. We also demonstrate that a signal sequence accounts for passage of polypeptides into the ER lumen, while only a small portion of the transit peptide-domain appears absolutely necessary to mediate transport across the double envelope membrane of the plastid. Furthermore, these studies raise significant questions with respect to the way in which polypeptides traverse the second most distal of the four membranes that delineate plastids, the function of the carboxy terminal region of the transit peptide and the role of the CER in polypeptide targeting.

This work was supported in part by grants from the Deutsche Forschungsgemeinschaft (DFG) to P.G.K. (SFB 189 and Transregio-SFB TR1).

References

Allison, D. S. and Schatz, G. (1986). Artificial mitochondrial pre-sequences. *Proc. Natl. Acad. Sci. USA* **83**, 9011-9015.

Andersen, R. A., Morton, S. L. and Sexton, J. P. (1997). CCMP-Provasoli-Guillard National center for culture of marine phytoplankton. *J. Phycol. (Suppl.)* **33**, 6.

Apt, K. E., Bhaya, D. and Grossman, A. R. (1994). Characterization of the genes encoding the light-harvesting proteins in diatoms: the biogenesis of the fucoxanthin chlorophyll a/c protein complex. *J. Appl. Phycol.* **6**, 225-230.

Apt, K. E., Clendennen, S. K., Powers, D. and Grossman, A. R. (1995). The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*. *Mol. Gen. Genet.* **246**, 455-464.

Apt, K. E., Sukenik, A. and Grossman, A. R. (1995). The ER chaperone BiP from the diatom *Phaeodactylum*. *Plant Physiol.* **109**, 339.

Apt, K. E., Kroth-Pancic, P. G. and Grossman, A. R. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* **252**, 572-579.

Baker, A. and Schatz, G. (1987). Sequences from a prokaryotic genome or the mouse dehydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc. Natl. Acad. Sci. USA* **84**, 3117-3121.

Bhaya, D. and Grossman, A. R. (1991). Targeting proteins to diatom plastids

involves transport through an endoplasmic reticulum. *Mol. Gen. Genet*. **229**, 400-404.

Borowitzka, M. A. and Volcani, B. E. (1978). The polymorphic diatom *Phaeodactylum tricornutum*: ultrastructure of its morphotypes. *J. Phycol*. **14**, 10-21.

Borowitzka, M. A., Chiappino, M. L. and Volcani, B. E. (1977). Ultrastructure of a chain-forming diatom *Phaeodactylum tricornutum*. *J. Phycol*. **13**, 162-170.

Bouck, G. B. (1965). Fine structure and organelle associations in brown algae. *J. Cell Biol*. **26**, 523-537.

Cavalier-Smith, T. (1986). The Kingdom Chromista: Origin and Systematics. *Progr. Phyc. Res.* **4**, 309-437.

Cavalier-Smith, T. (2000). Membrane heredity and early chloroplast evolution. *Trends Plant Sci*. **5**, 174-182.

Chen, X. J. and Schnell, D. J. (1999). Protein import into chloroplasts. *Trends Cell Biol*. **9**, 222-227.

Crawford, R. M. (1973). The protoplasmic ultrastructure of the vegetative cell of *Melosira varians* C.A. Agardh. *J. Phycol*. **9**, 50-61.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat*. **154**, S164-S177.

Delwiche, C. F. and Palmer, J. D. (1997). The origin of plastids and their spread via secondary symbiosis. *Plant. Syst. Evol*. **11**, 53-86.

Gibbs, S. P. (1978). The chlorochloroplasts plasts of *Euglena* may have evolved from symbiotic green alga. *Can. J. Bot*. **56**, 2883-2889.

Gibbs, S. P. (1979). The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. *J. Cell Sci*. **35**, 253-266.

Gibbs, S. P. (1981). The chloroplast endoplasmic reticulum: structure, function and evolutionary significance. *Int. Rev. Cytol*. **72**, 49-99.

Grossman, A. R., Manodori, A. and Snyder, D. (1990). Light-harvesting proteins of diatoms: their relationship to the chlorophyll *a/b* binding proteins of higher plants and their mode of transport into plastids. *Mol. Gen. Genet*. **224**, 91-100.

Grossman, A. R., Bhaya, D., Apt, K. E. and Kehoe, D. M. (1995). Light-harvesting complexes in oxygenic photosynthesis: diversity, control and evolution. *Annu. Rev. Genet*. **29**, 231-299.

Heins, L., Collinson, I. and Soll, J. (1998). The protein translocation apparatus of chloroplast envelopes. *Trends Plant Sci*. **3**, 56-61.

Ishida, K., Cavalier-Smith, T. and Green, B. R. (2000). Endomembrane structure and the chloroplast protein targeting pathway in *Heterosigma akashiwo* (Raphidophyceae, Chromista). *J. Phycol*. **36**, 1135-1144.

Kroth, P. G. (2002). Protein transport into secondary plastids and the evolution of primary and secondary plastids. *Int. Rev. Cytol*. **221**, (in press).

Kroth, P. and Strotmann, H. (1999). Diatom plastids: secondary endocytobiosis, plastid genome and protein import. *Physiol. Plant* **107**, 136-141.

Lang, M., Apt, K. E. and Kroth, P. G. (1998). Protein transport into 'complex' diatom plastids utilizes two different targeting signals. *J. Biol. Chem*. **273**, 30973-30978.

Liaud, M.-F., Brundt, U., Scherzinger, M. and Cerff, R. (1997). Evolutionary origin of cryptomonad microalgae: two novel chloroplast/cytosol-specific GAPDH genes as potential markers of ancestral endosymbiont and host cell components. *J. Mol. Evol*. **44**, S28-S37.

Martin, W. and Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiol*. **118**, 9-17.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M. and Kowallik, K. V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162-165.

McFadden, G. I. (1999). Plastids and protein targeting. *J. Euk. Microbiol*. **46**, 339-346.

Moreira, D., le Guyader, H. and Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69-72.

Pancic, P. G. and Strotmann, H. (1993).: Structure of the nuclear encoded gamma subunit of CF_0CF_1 of the diatom *Odontella sinensis* including its presequence. *FEBS Lett*. **320**, 61-66.

Pont-Kingdon, G. (1997). Creation of chimeric junctions, deletions, and insertions by PCR. *Methods Mol. Biol.* **67**, 167-172.

Sato, T. (1968). A modified method for lead staining of thin sections. *J. Electron Microsc.* **17**, 158-159.

Smith-Johannsen, H. and Gibbs, S. P. (1972). Effects of chloramphenicol on chloroplast and mitochondrial ultrastructure in *Ochromonas danica*. *J. Cell Biol*. **52**, 598-614.

Starr, R. C. and Zeikus, J. A. (1993). UTEX-The culture collection of algae at the University of Texas at Austin. *J. Phycol*. *(Suppl.)* **29**, 93.

Sulli, C., Fang, Z. W., Muchhal, U. S. 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.

Tréguer, P., Nelson, D. M., van Bennekom, A. J., DeMaster, D. J., Leynaert, A. and Quéquiner, B. (1995). The silica balance in the world ocean: a reestimate. *Science* **268**, 375-379.

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.

von Heijne, G., Steppuhn, J. and Hermann, R. G. (1989). Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem*. **180**, 535-545.

Walter, P. and Johnson, A. E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol*. **10**, 87-119.

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.

Werner, D. (1977). Introduction with an note on taxonomy. In *The Diatoms* (ed. D. Werner), pp. 1-17. Oxford: Blackwell Science.

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R. and Apt, K. E. (2000). Additional selectable marker and reporter genes for use in transforming the diatom *Phaeodactylum tricornutum*. *J. Phycol*. **36**, 379-386.

Zaslavskaia, L. A., Lippmeier, J. C., Shih, C., Ehrhardt, D., Grossman, A. R. and Apt, K. E. (2001). Trophic conversion of an obligate photoautotrophic organism through metabolic engeneering. *Science* **292**, 2073-2075.