Plant Mol Biol (2007) 64:519–530
DOI 10.1007/s11103-007-9171-x

# Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif

**Ansgar Gruber · Sascha Vugrinec · Franziska Hempel · Sven B. Gould · Uwe-G. Maier · Peter G. Kroth**

Received: 7 December 2006 / Accepted: 30 March 2007 / Published online: 5 May 2007


**Abstract** Plastids of diatoms and related algae evolved by secondary endocytobiosis, the uptake of a eukaryotic alga into a eukaryotic host cell and its subsequent reduction into an organelle. As a result diatom plastids are surrounded by four membranes. Protein targeting of nucleus encoded plastid proteins across these membranes depends on N-terminal bipartite presequences consisting of a signal and a transit peptide-like domain. Diatoms and cryptophytes share a conserved amino acid motif of unknown function at the cleavage site of the signal peptides (ASAFAP), which is particularly important for successful plastid targeting. Screening genomic databases we found that in rare cases the very conserved phenylalanine within the motif may be replaced by tryptophan, tyrosine or leucine. To test such unusual presequences for functionality and to better understand the role of the motif and putative receptor proteins involved in targeting, we constructed presequence:GFP fusion proteins with or without modifications of the ''ASAFAP''-motif and expressed them in the diatom *Phaeodactylum tricornutum*. In this comprehensive mutational analysis we found that only the aromatic amino acids phenylalanine, tryptophan, tyrosine and the bulky amino acid leucine at the +1 position of the predicted signal peptidase cleavage site allow plastid import, as expected from the sequence comparison of native plastid targeting presequences of *P. tricornutum* and the cryptophyte *Guillardia theta*. Deletions within the signal peptide domains also impaired plastid import, showing that the presence of F at the N-terminus of the transit peptide together with a cleavable signal peptide is crucial for plastid import.

**Keywords** Chloroplast · Diatom · Evolution · Import · Presequence

A. Gruber and S. Vugrinec contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-007-9171-x) contains supplementary material, which is available to authorized users.

A. Gruber · S. Vugrinec · P. G. Kroth (✉)
Plant Ecophysiology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany
e-mail: peter.kroth@uni-konstanz.de

F. Hempel · S. B. Gould · Uwe-G.Maier
Cell Biology, Philipps-University Marburg, Karl-von-Frisch Straße 8, 35042 Marburg, Germany

*Present Address:*
S. B. Gould
School of Botany, University of Melbourne, Melbourne, VIC 3010, Australia

## Introduction

According to the actual view, all plastids can be traced back to an endosymbiotic event in which a cyanobacterium was taken up by a eukaryotic cell, followed by the reduction of the endosymbiont to an organelle. The resulting primary plastids are monophyletic and are found in glaucophytes, rhodophytes, chlorophytes and land vascular plants (Martin et al. 1998; Moreira et al. 2000; Rodriguez-Ezpeleta et al. 2005). Diatoms and other groups of algae possess secondary plastids which originated from a secondary endocytobiosis event: the uptake of a eukaryotic alga possessing primary plastids into a heterotrophic host cell. This endosymbiotic alga again was subsequently reduced to a plastid. Secondary plastids are surrounded by either three or four membranes and thus are also known as complex plastids (Cavalier-Smith 1999, 2000; McFadden 2001). Secondary endocytobiosis was a key event during

the evolution of a variety of organisms and was found to have occurred at least twice, as some complex plastids have a green algal origin while others are related to red algae (Cavalier-Smith 1999, 2000). There is increasing evidence that the secondary plastids of the red algal lineage originate from a single endosymbiotic event and that the resulting chromalveolates (including heterokonts, cryptophytes, haptophytes, apicomplexa and dinoflagellates) might be monophyletic (Cavalier-Smith 1999; Harper et al. 2005). While cryptophytes still possess a remnant nucleus of the endosymbiont, the nucleomorph, which is located in the periplastidic space between the second and third envelope membrane, in heterokonts (including diatoms) the reduction of the endosymbiont included the loss of the endosymbiont's nucleus, the mitochondria and all other cytoplasmatic components (Keeling 2004). In apicomplexean parasites (like *Plasmodium falciparum*, the causative organism of malaria) also the plastid itself is highly reduced (with respect to genome size and endomembranes) down to the colourless and non-photosynthetic apicoplast (Waller and McFadden 2004).

During the reduction of the primary and secondary endosymbiotic cells, most of the genes of the endosymbiont were either lost, replaced by genes of the host or transferred to the nucleus of the host cell (Delwiche 1999; Martin and Herrmann 1998; Timmis et al. 2005). Therefore an efficient plastid protein import system had to be established in order to provide the organelles with plastid proteins now encoded in the nucleus (Ishida 2005; Kroth 2002). This must have been quite a challenge since at least 1240 plastid proteins were experimentally identified in the higher plant *Arabidopsis thaliana* (Heazlewood et al. 2005), while the plastid proteome of *A. thaliana* in total was estimated to consist of about 2,700 different proteins (Millar et al. 2006). Protein targeting across the two envelope membranes of the primary plastids of land plants is well characterised and is mainly based on posttranslational import by two protein translocator complexes called translocator of the outer/inner chloroplast envelope membrane (Toc and Tic) and a subsequent cleavage of the N-terminal targeting signal called transit peptide (Soll and Schleiff 2004).

In cryptophytes and diatoms there are two additional membranes around the plastids, the outermost being studded with ribosomes and continuous with the endoplasmic reticulum (ER) (Gibbs 1981). The plastid genomes of the diatom *Phaeodactylum tricornutum* and the cryptophyte *Guillardia theta* contain only 162 and 177 genes (Douglas and Penny 1999; Oudot-Le Secq et al. 2007), however, a plastid proteome size similar to that of higher plants must be assumed because photosynthesis is a rather complex process. Plastid protein import is therefore an important process for diatoms and cryptophytes, but the mode of protein translocation into these complex plastids derived from red algae is still mysterious (Kilian and Kroth 2003). Presequences of nucleus encoded plastid proteins consist of a signal peptide followed by a transit peptide-like domain (Pancic and Strotmann 1993). The functionality of both domains was proven individually *in vitro* in heterologous import systems (Bhaya and Grossman 1991; Chaal and Green 2005; Ishida et al. 2000; Lang et al. 1998; Nassoury et al. 2003; Wastl and Maier 2000), and previous studies in the diatom *Phaeodactylum tricornutum* demonstrated the *in vivo* functionality of native plastid presequences:GFP fusion proteins (Apt et al. 2002). Interestingly, also heterologous presequences from the diatom *Odontella sinensis* (Kilian and Kroth 2005; Kroth et al. 2005) or from the dinoflagellate *Symbiodinium* sp. (Lang 2000) and the cryptophyte *G. theta* (Gould et al. 2006a) were able to direct GFP into the plastid of *P. tricornutum*, indicating similarities between the plastid protein import machineries in cryptophytes, dinoflagellates and diatoms.

Another striking similarity of cryptophytes and diatoms is the presence of a conserved amino acid motif at the signal peptide's predicted cleavage site (ASAFAP) in both algal groups (Gould et al. 2006a; Kilian and Kroth 2005; Kroth 2002). Unlike most other import systems based on cleavable presequences, here the presence of a single amino acid is most crucial for plastid import. Also surprisingly large parts of the C-terminus of the transit peptide-like domain can be deleted without affecting protein transport into diatom plastids *in vivo* (Apt et al. 2002), while the exchange of phenylalanine within the ''ASAFAP''-motif may block protein import completely (Kilian and Kroth 2005). Although this phenylalanine is highly conserved, recent large scale sequencing projects on diatoms and cryptophytes revealed a few presequences that contain other aromatic amino acids (tryptophan, tyrosine) or leucine instead.

To evaluate the necessity of individual amino acids within the presequence and to collect information about possible receptor proteins that recognise the presequences, we tested these presequences by in vivo experiments and modified existing presequences by site directed mutagenesis. We demonstrate that most modifications concerning the phenylalanine within the ''ASAFAP''-motif block plastid import of the respective fusion protein, while a few other substitutions in the same position allow plastid import.

## Materials and methods

### Sequence analysis

We screened sequences from a *Guillardia theta* EST project (Gould et al. 2006a, b) and from the current US Department of Energy Joint Genome Institute (JGI, http://

www.jgi.doe.gov/) diatom genome sequencing projects for the diatoms *Thalassiosira pseudonana* (http://genome.jgi-psf.org/Thaps3/Thaps3.home.html) (Armbrust et al. 2004) and *Phaeodactylum tricornutum* (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) (Bowler et al., in preparation) as well as publicly available databases of sequences from secondary algae for sequences with homology to plastid proteins using the BLAST algorithm (Altschul et al. 1997). Resulting hits were screened for the presence of signal peptides using the program SignalP (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen et al. 2004). For cleavage site predictions the results of SignalP's Neuronal networks (NN) (Nielsen et al. 1997b) or Hidden Markov Models (HMM) (Nielsen and Krogh 1998) were used; for prediction of chloroplast transit peptide-like domains, the programs ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) (Emanuelsson et al. 1999) and TargetP (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al. 2000) were utilised. The transit peptide-like domains of bipartite plastid targeting sequences often attain poor prediction scores, so we used the NCBI (http://www.ncbi.nlm.nih.gov/) Conserved Domain Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2005) to identify N-terminal extensions from the conserved regions of the respective protein. If a distance of at least 10 amino acids between the predicted cleavage site of the signal peptide and the region of high homology to respective proteins of other organisms was found, also a weakly predicted transit peptide-like domain was accepted. Sequence logos (Schneider and Stephens 1990) were prepared using the WebLogo server (http://weblogo.berkeley.edu/) (Crooks et al. 2004) to illustrate the predictions of the different algorithms with predictions combining computation and manual correction.

## Plasmid constructs

Standard cloning procedures were applied (Sambrook et al. 1989). Polymerase chain reaction (PCR) was performed with a Master Cycler Gradient (Eppendorf, Hamburg, Germany) using recombinant Pfu polymerase (Fermentas, GmbH, St. Leon-Rot, Germany) according to the manufacturer's instructions. All presequences used in this work are based on cDNAs derived from *Phaeodactylum tricornutum* or from *Guillardia theta*. To produce the *G. theta* GtPGK:GFP (GenBank AM413041) and the *P. tricornutum* PtFBAC1:GFP (GenBank AY191866) constructs, GFP fusions were inserted into the EcoRI and HindIII restriction sites of the *Phaeodactylum tricornutum* transformation vector pPha-T1 (GenBank AF219942, Zaslavskaia et al. 2000). Unmodified presequences were amplified by PCR, including 5–8 base pairs upstream of the start codon to facilitate initiation of translation (Kozak 1987). Homologous primer pairs contained EcoRI and NcoI restriction sites within the upstream or downstream primers, respectively. Fusions of the plastid preprotein presequences to the gene encoding the enhanced green fluorescent protein (GFP) were performed by using an NcoI restriction site containing the start codon of the GFP gene (BD Bioscience, Palo Alto, CA, USA). For the PtOEE1:GFP (GenBank AY191862, Protein ID 20331, annotated in the *P. tricornutum* genome database) fusion protein the downstream primer for presequence amplification contained the restriction sites XbaI and XhoI leading to the derived artificial amino acid sequence ''SRMLE'' (indicated in Fig. 2). Here, the presequence was fused to the GFP gene via an XhoI restriction site and the GFP fusion was inserted into the EcoRI and HindIII restriction sites of pPha-T1. Construction of the GFP fusion proteins PtOEE1:GFP and PtFBAC1:GFP has been described in more detail previously (Apt et al. 2002; Kilian and Kroth 2004, 2005). For the construction of the PtHLIP2:GFP (Protein ID: 55112), PtFBPC4:GFP (Protein ID: 54279) and PtFSA:GFP (Protein ID: 20779) fusion proteins GFP has been amplified in a first step, adding the recognition site for StuI upstream of the start codon ATG, which allowed in frame cutting. The shuttle vector pPha-T1 was linearised using EcoRV and the modified GFP fragment was ligated into the plasmid in the orientation of the fcpA promoter, resulting in the plasmid pPha-T1-GFP. The presequences were amplified using unmodified Primers. After digesting pPha-T1-GFP with StuI the presequence amplification products were ligated into the plasmid upstream of and in frame with GFP. All constructs were sequenced from their 5′ end, to ensure correct cloning.

Site directed point mutagenesis was performed with the QickChange mutagenesis Kit (Stratagene, La Jolla, CA, USA) according to the protocol supplied by the manufacturer. The artificial sequence information was modified or inserted according to the codon usage in *P. tricornutum* (Montsant et al. 2005), the most common codons for the modified amino acids were used. Plasmids have been sequenced to control weather the introduced modifications were incorporated properly.

## Culture conditions

*Phaeodactylum tricornutum* Bohlin (University of Texas Culture Collection, strain 646) was cultivated in Provasoli's enriched seawater (Starr and Zeikus 1993) using ''Tropic Marin'' (Dr. Biener GmbH, Wartenberg, Germany) salt (16.6 g $l^{-1}$), 50% concentration compared to natural seawater. Cells were grown in liquid culture in flasks under rigorous shaking (120 rpm) at 22°C with continuous illumination at 35 µmol·photons·$m^{-2}$ $s^{-1}$. Solid media contained 1.2% (w/v) Bacto Agar (BD, Sparks, MD, USA).

### Nuclear transformation

Nuclear transformation of *Phaeodactylum tricornutum* has been performed using a Bio-Rad Biolistic PDS-1000/He Particle Delivery System (Bio-Rad, Hercules, CA, USA) fitted with 1350 psi rupture disks as described previously (Apt et al. 1996) and recently in more detail (Kroth 2007). For the selection and cultivation of *P. tricornutum* transformants 75 μg $ml^{-1}$ Zeocin (Invitrogen, Carlsbad, CA, USA) was added to the solid medium.

### Microscopy

Cells were observed using an Olympus BX51 epifluorescence microscope equipped with a Nikon DXM1200 digital camera system (Olympus Europe, Hamburg, Germany). Nomarski's differential interference contrast illumination was used to view transmitted light images. Chlorophyll autofluorescence and green GFP fluorescence of the transformants have been dissected using the mirror unit U-MWSG2 (Olympus) and the filter set 41020 (Chroma Technology Corp, Rockingham, VT, USA), respectively. Multichannel fluorescence pictures were taken and assembled with the software LUCIA (Nikon GmbH, Düsseldorf, Germany). The micrographs were size calibrated using a stage micrometer.

## Results

### Sequence analysis

In earlier work it was demonstrated that most nuclear encoded plastid preproteins of diatoms and cryptophytes contain a phenylalanine in the region of the signal peptide cleavage site (Armbrust et al. 2004; Gould et al. 2006a; Kilian and Kroth 2005). We analysed genes of further plastid preproteins by screening the whole genome databases of *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* as well as EST sequences of *Guillardia theta* and public databases for plastid preproteins of other related algae. Although most gene products assigned as plastid proteins contain the respective phenylalanine, in some cases either a tryptophan, a leucine or a tyrosine are present in the expected position instead. To access the frequency of such unusual presequences and to evaluate alternative prediction models we performed a genome wide comparison of plastid presequences from *Phaeodactylum tricornutum*. Among 81 manually curated plastid gene models within the first release of the genome v1.0 (http://genome.jgi-psf.org/Phatr1/Phatr1.home.html) (Bowler et al., in preparation), 72 contain a phenylalanine at the signal peptide cleavage site. We found six sequences containing tryptophan, two sequences containing leucine and one sequence containing tyrosine at the signal peptide cleavage site (supplementary Fig. 6).

The predicted signal peptide cleavage sites may vary depending on the calculation method. Two options are available for SignalP (Bendtsen et al. 2004), prediction by NN (Nielsen et al. 1997b) or by HMM (Nielsen and Krogh 1998). The prediction was identical in 68 of the 81 tested sequences, and in 62 of these cases the predicted cleavage site coincided with an ''ASAFAP''-motif. In 10 cases the predictions differed between the models, but one of the predicted cleavage sites coincided with an ''ASAFAP''-motif (6 × NN, 4 × HMM), and in three cases both models predicted different cleavage sites without ''ASAFAP''-motif (supplementary Fig. 6). Manual analysis of these sequences revealed an ''ASAFAP''-motif (often reduced to ''AF'') in proximity to the predicted cleavage site. Sequence logos (Schneider and Stephens 1990) created from these data sets reveal a conserved motif flanking the predicted cleavage site (Fig. 1). Sequence conservation and the resulting conserved sequences are displayed at the same time when the sequence is printed as a stack of letters, with the height of a stack representing the sequence conservation at that position. Sequence conservation is higher the fewer the number of residues at one position is, resulting in a higher information content measured as bits. The height of an amino acid letter in the stack is proportional to its frequency, with the most frequent residue printed on top of the stack (Schneider and Stephens 1990). Different sequence logos have been prepared, relying on different prediction algorithms (Fig. 1 and supplementary Fig. 6).

For the NN and HMM prediction sequence logos (Fig. 1, upper left and upper right) the respective predictions were used as indicated in supplementary Fig. 6. For the ''highest prediction'' sequence logo (Fig. 1, lower left) the model with the highest prediction score (Ymax for NN versus Cmax from NN) was used, as printed in bold in supplementary Fig. 6. A sequence logo combining automated prediction with manual corrections additionally considering the presence of an ''ASAFAP''-motif (Fig. 1, lower right) was prepared using the automated prediction if it was identical between NN and HMM and coincided with an ''ASAFAP''-motif (supplementary Fig. 6a). If the predictions differed between the models, predictions were chosen if they coincided with an ''ASAFAP''-motif (supplementary Fig. 6b), or an ''ASAFAP''-motif in proximity to an automatically predicted cleavage site was chosen if there was no exact coincidence of an automatically predicted cleavage site with an ''ASAFAP''-motif (supplementary Fig. 6c). The cleavage site motifs used for the ''manual prediction'' sequence logo (Fig. 1, lower right) are indicated in grey in supplementary Fig. 6. Sequence conservation and proportion of phenylalanine is slightly

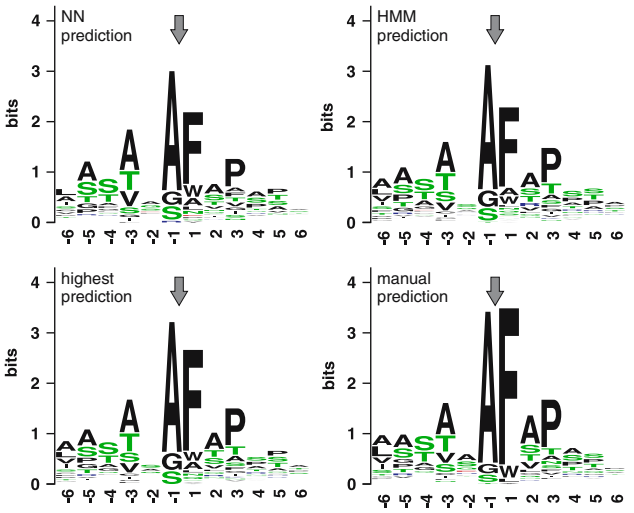


**Fig. 1** Sequence logos constructed from 81 manually curated plastid gene models (see also supplementary Fig. 6) within the *Phaeodactylum tricornutum* genome. Predictions from Signal P's Neuronal networks (NN, upper left) and Hidden Markov models (HMM, upper right) can be compared. In addition a combined sequence logo using the highest prediction score (Ymax for NN versus Cmax from NN, lower left) has been prepared. A sequence logo using manual predictions, where the automated outputs of NN and HMM have been corrected with respect to the presence of an ''ASAFAP''-motif (see supplementary Fig. 6 for the exact corrections applied) shows the highest sequence conservation surrounding the signal peptide cleavage site (lower right). Black: hydrophobic residues (ACFGILMPVWY), green: hydrophilic residues (NQST), blue: basic residues (HKR) red: acidic residues (DE)

higher at the +1 position of the predicted cleavage site in the HMM prediction compared to the NN prediction. When combining both models, sequence conservation of the –1 and +1 position of the predicted cleavage site improves, but the highest conservation is obtained when the automated predictions of NN or HMM are corrected manually depending on the presence of an ''ASAFAP''-motif close to the automatically predicted cleavage site (Fig. 1 and supplementary Fig. 6).

## Native plastid targeting sequences

To test the functionality of different plastid protein sequences we fused the respective gene fragments encoding presequences of interest to the GFP gene and expressed the fusion proteins in *Phaeodactylum tricornutum* (Fig. 2). We found that the fusion proteins were correctly imported into the plastids and that the GFP fluorescence colocalised with the chlorophyll autofluorescence of the plastid. The *P. tricornutum* (Pt) PtOEE1 and the PtFBAC1 presequences, both containing classical ''ASAFAP''-motifs, were imported into the plastids as expected (Fig. 3b, c). Also the PtHLIP2 and the PtFSA presequences, containing an ''AW''-cleavage site motif instead, led to GFP fluorescence in the plastids of transformed cells (Fig. 3a). Similarly, the presequence of the heterologous *Guillardia theta* (Gt) GtPGK protein fused to GFP (tyrosine instead of phenylalanine) was imported correctly into *P. tricornutum* plastids (Fig. 3d). The PtFBPC4 presequence:GFP fusion construct containing an ''AW''-cleavage site motif was the only exception and gave ambiguous results. In some transformant cell lines GFP was fluorescing inside the plastids, while in others GFP fluorescence also appeared outside of the plastids. The ''blob''-like structures described below were never observed in these cell lines (data not shown).

## Mutations of the signal peptide's cleavage site

The presequence of the *Phaeodactylum tricornutum* Oxygen evolving enhancer 1 (PtOEE1) protein has previously been characterised intensively (Kilian and Kroth 2004, 2005). This protein is normally targeted into the thylakoids (Ammon and Kroth, unpublished); for a better visualisation of GFP fluorescence, in all of the following constructs the third targeting domain responsible for thylakoid targeting has been deleted (Kilian and Kroth 2005). To confine crucial features of presequences for plastid import we introduced various point mutations into the presequence of the PtOEE1 presequence:GFP fusion protein (Fig. 4). In *P. tricornutum* transformants expressing the wild type presequence:GFP fusion protein PtOEE1:GFP, the GFP accumulated as expected in the chloroplast stroma (Fig. 3b). Deletion of the phenylalanine at the N-terminus of the transit peptide-like domain in the fusion protein PtOEE1Δ18F:GFP lead to a phenotype previously described as ''blob''-like structure (BLS), (Kilian and Kroth 2005), representing an accumulation of GFP in a small reticular structure tightly associated to the plastid but clearly outside the stroma (Fig. 3b). There are several indications that these structures accumulate between the plastid bounding membranes (Kilian and Kroth 2005).

By site directed mutagenesis we replaced the phenylalanine by other aromatic residues, like tyrosine, tryptophane and histidine and expressed the constructs in *P. tricornutum*. Substitution of phenylalanine by tyrosine and tryptophan (PtOEE1F18Y:GFP, PtOEE1F18W:GFP) resulted in functional targeting of GFP into the plastids (Fig. 3b), while a replacement by histidine (PtOEE1F18H:GFP) led to the BLS phenotype (Fig. 3b). Phenylalanine, tyrosine and tryptophan are large and hydrophobic amino acids, so we tested whether it would be sufficient to introduce other large and hydrophobic residues instead of phenylalanine. We inserted leucine, isoleucine and methionine and found that only leucine (PtOEE1F18L:GFP) at this position is capable of driving protein import (Fig. 3b), while PtOEE1F18I:GFP and PtOEE1F18M:GFP again led to the BLS phenotype

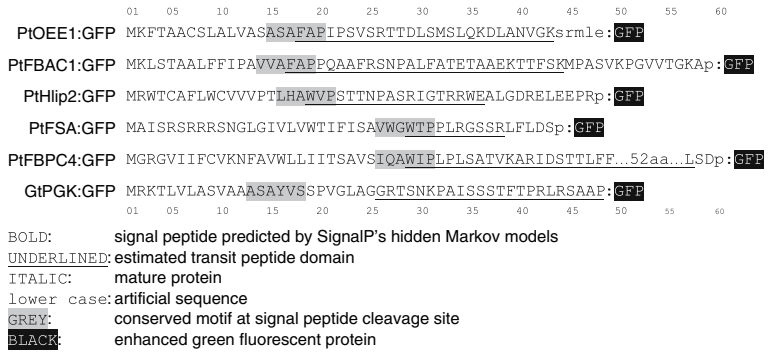


**Fig. 2** Unmodified presequences fused to enhanced green fluorescent protein (GFP). PtOEE1 (oxygen evolving enhancer protein 1), PtFBAC1 (fructose-1,6-bisphosphate aldolase), PtHlip2 (high light induced protein 2), PtFSA (fructose-6-phosphate-aldolase) and PtFBPC4 (fructose-bisphosphatase) presequence domains are from *Phaeodactylum tricornutum* (Pt). The GtPGK (phosphoglycerate kinase) presequence domain is from *Guillardia theta* (Gt), this presequence is an example where the conserved ''ASAFAP''-motif does not coincide with the signal peptide's predicted cleavage site. All fusion proteins result in plastid import of GFP when expressed in *P. tricornutum*

(data not shown). In *P. tricornutum* cells with a strong expression of PtOEE1F18L:GFP we also found weak labeling of the cytosol, which was never observed in wild type PtOEE1:GFP or in the PtOEE1F18Y:GFP and PtOEE1F18W:GFP transformants. This may indicate that import of PtOEE1F18L:GFP into the chloroplast endoplasmic reticulum (CER) is less efficient than the import of fusion proteins with the aromatic residues phenylalanine, tyrosine or tryptophan at the N-terminus of the transit peptide-like domain. More likely this phenotype is an overexpression artefact, since several other transformed cell lines showed fairly GFP labelled plastids. We repeated the experiment by re-sequencing the PtOEE1F18L:GFP plasmid construct and re-transforming *P. tricornutum* with the plasmid to ensure this particular result. We modified the presequence of the *P. tricornutum* fructose bisphosphatase (PtFBAC1) in a similar way, changing phenylalanine to leucine, and obtained similar results: the resulting fusion protein PtFBAC1F17L:GFP again was imported into the plastid (Fig. 3c). A replacement of phenylalanine by charged amino acids like arginine (PtOEE1F18R:GFP) and glutamate (PtOEE1F18E:GFP) and by the small residue glycine (PtOEE1F18G:GFP) did not result in plastid import and transformants showed the BLS phenotype (data not shown). To test the importance of the exact position of the phenylalanine we exchanged the amino acids F and A flanking the signal peptide's cleavage site, the resulting PtOEE1A17F+F18A:GFP construct led to the BLS phenotype (Fig. 3b).

### Mutations of the transit peptide-like domain

We inserted mutations in the transit peptide-like domain of the motif to assess the importance of these residues for successful plastid import. The deletion mutants PtOEE1-Δ19A:GFP and PtOEE1Δ20P:GFP allowed plastid import of the GFP (sequences listed in Fig. 4), furthermore we were able to exchange proline by alanine (PtOEE1-P20A:GFP) without affecting plastid import (data not shown). Replacement of all alanine residues within the ''ASAFAP''-motif by serine and glycine did not affect the plastid protein import, as the fusion proteins PtOEE1A(15–19)S:GFP and PtOEE1A(15–19)G:GFP are targeted into the plastids (data not shown).

### Mutations of the signal peptide domain

In contrast, deletions of alanine and serine within the signal peptide part of the ''ASAFAP''-motif blocked plastid protein targeting. The BLS phenotype was observed in transformants expressing the fusion proteins PtOEE1Δ15A:GFP, PtOEE1Δ16S:GFP, PtOEE1Δ17A:GFP and PtOEE1-Δ17A+A19S:GFP (PtOEE1Δ17A:GFP shown as example in Fig. 3). Exchange of the serine to alanine or cysteine led to correct plastid import, transformants expressing the PtOEE1S16A:GFP and PtOEE1S16C:GFP showed GFP fluorescence within the plastids (data not shown). Also the exchange of alanine to serine preceding the tyrosine in the *G. theta* GtPGK presequence did not affect plastid import of the fusion protein (Fig. 3d). The absence of serine in the signal peptide of the wild type *P. tricornutum* PtFBAC1 presequence (Fig. 4b) shows that the presence of serine in the signal peptide is not required for successful plastid targeting, although serine is commonly found within the ''ASAFAP''-motif and within plastid protein signal peptides (Figure 1 and supplementary Fig. 6).

## Discussion

The development of protein targeting into the secondary plastids of diatoms and cryptophytes was a prerequisite for

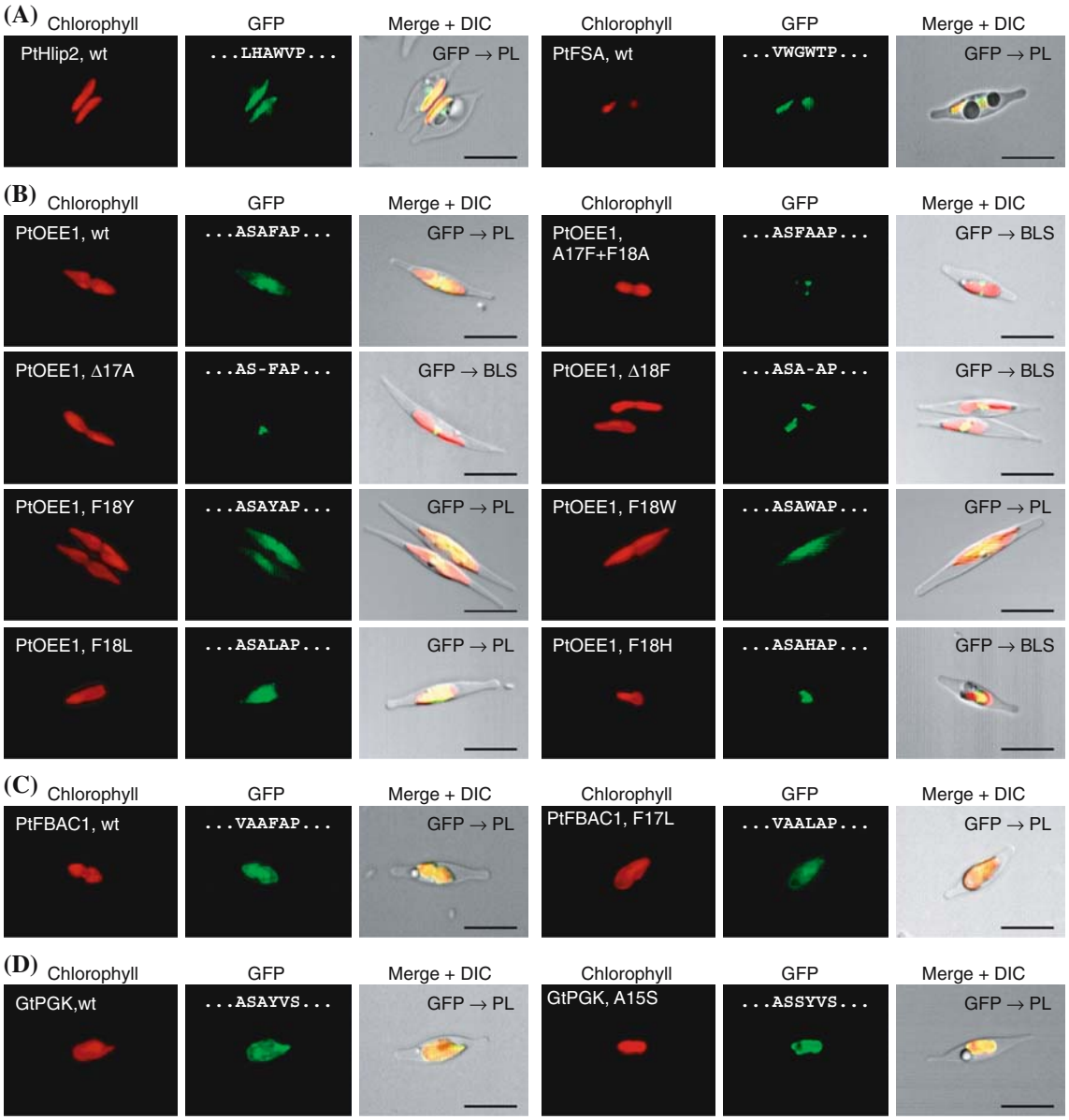


**Fig. 3** Localisation of the presequence:GFP fusion proteins after expression in *Phaeodactylum tricornutum*. Wild type (wt) or mutated presequences (see also Fig. 4) of plastid proteins lead to import of GFP into the plastid (PL) or into ''blob''-like structures (BLS). **(A)** Wild type presequences of PtHlip2 (high light induced protein 2) and PtFSA (fructose-6-phosphate-aldolase) from *P. tricornutum*. **(B)** Wild type and modified presequences of the PtOEE1 (oxygen evolving enhancer protein 1) from *P. tricornutum*. **(C)** Wild type and modified presequence of the PtFBAC1 (fructose-1,6-bisphosphate aldolase) from *P. tricornutum*. **(D)** Wild type and modified presequence of the GtPGK (phosphoglycerate kinase) from *Guillardia theta*. Red chlorophyll autofluorescence, green GFP fluorescence and a merge of Chlorophyll and GFP fluorescences with Normarski differential interference contrast (DIC) images are shown from left to right, scale bars represent 10 μm

the successful establishment of secondary endosymbioses, because it allowed gene transfer from the endosymbiont to the nucleus of the host cell and a transport of the respective gene products into the endosymbiont/organelle (Cavalier-Smith 1999, 2003). Genes that shifted from the endosymbiont's nucleus to the nucleus of the host cell likely already contained transit sequences and needed a signal sequence for completion, while genes that shifted directly from the plastid genome to the nucleus needed the whole targeting domain (Kilian and Kroth 2004). In diatoms these presequences consist of a signal peptide domain and a transit peptide-like domain (Pancic and Strotmann 1993), reflecting this evolutionary history.

The signal peptide domains and the transit peptide-like domains have been found to individually facilitate ER import and import into primary plastids, respectively, by

*in vitro* experiments (Bhaya and Grossman 1991; Lang et al. 1998). *In vivo* experiments showed that these bipartite presequences are sufficient for plastid import and that no other targeting signals are needed (Apt et al. 2002; Kilian and Kroth 2004, 2005). Interestingly although large parts of the C-terminus of the transit peptide-like domain may be deleted (Apt et al. 2002), plastid import is only possible if a conserved ''ASAFAP''-motif is present between the signal and the transit peptide-like domains (Kilian and Kroth 2005). Complete deletion of either the transit peptide-like domain or the phenylalanine within the ''ASAFAP''-motif lead to transport inhibition demonstrating that both elements are necessary. The very conserved phenylalanine within the ''ASAFAP''-motif has already been shown to be crucial for plastid targeting in a previous study (Kilian and Kroth 2005). Here we demonstrate that only a few structurally similar amino acids may replace this particular amino acid, while in all other cases exchanges of phenylalanine lead to blocked import. All other amino acids of the ''ASAFAP''-motif may be replaced by glycine, alanine, serine or cysteine without affecting import (Fig. 5). Interestingly deletions in the signal-peptide part of the motif could block plastid import, while exchanges at the same positions allowed plastid import. Possibly due to the shorter distance to the N-terminus in these cases the prediction of the cleavage site shifted, which might explain why the respective proteins are no longer imported. In these cases the phenylalanine is predicted to be cleaved off together with the signal peptide (Fig. 4, PtOEE1Δ15A, PtOEE1Δ16S, PtOEE1Δ17A). However, in some cases the phenylalanine (or the compensating tryptophan) is also predicted to be cleaved off, but the mutated presequence:GFP fusion proteins are imported into the plastid (Fig. 4, PtOEE1F18W, PtOEE1Δ20P, PtOEE1P20A). Probably because the overall length of the signal peptide is not affected, cleavage in these cases takes place as usual, regardless of the prediction.

The following requirements for preprotein import into complex diatom plastids can be deduced from this and from the former studies: (i) The presence of a cleavable signal peptide. (ii) The presence of predominately phenylalanine, sometimes tryptophan, rarely tyrosine or leucine in the +1 position of the signal peptide cleavage site, often followed by ''AP'' or a transit peptide domain. The ''ASAFAP''-motif fulfills these requirements, as the pre-cleavage site part of the motif can be explained by the ''(–3, –1) rule'' (von Heijne 1983) for cleavable signal peptides and the post-cleavage site part of the motif reflects the second requirement (presence of phenylalanine, tryptophan, tyrosine or leucine). The ''(–3, –1) rule'' is followed to a lesser extent in eukaryotic signal peptides compared to their prokaryotic counterparts (Nielsen et al. 1997a). Comparison of our sequence logos (Fig. 1) to sequence logos of prokaryotic and eukaryotic signal peptides (Nielsen et al. 1997a) shows that in *P. tricornutum* the conservation of signal peptides in the –3, –1 positions is higher than generally found in eukaryotes. This finding might reflect the fact that signal peptide cleavage is crucial in the process of plastid protein import into complex diatom plastids.

Generally the predicted signal peptide cleavage sites may vary depending on the calculation method. The possibility of miss-predictions complicates bioinformatic attempts to recognise plastid proteins. A hand selected sequence logo of plastid targeting signals revealed the presence of the conserved cleavage site motif in all tested sequences, but it was constructed from known plastid proteins only (Kilian and Kroth 2005). The sequence logo of a genome wide automated comparison of *Thalassiosira pseudonana* transit peptides also showed other amino acids than phenylalanine, tryptophan, tyrosine or leucine in the first position of the transit peptide-like domain, predicting alanine to be the second frequent amino acid in this position (Armbrust et al. 2004). This is contradictory to our finding that replacements of phenylalanine by structurally dissimilar amino acids like alanine lead to blocked plastid import. Here, only nine native plastid targeting sequences contained an ''ASAFAP''-motif without phenylalanine at the signal peptide cleavage site and until now we only observed native plastid presequences containing the structurally similar amino acids shown to functionally replace phenylalanine (tryptophan, tyrosine or leucine) in this position (supplementary Fig. 6).

We conclude that the occurrence of alanine as the second frequent N-terminal amino acid in a bioinformatic approach (Armbrust et al. 2004) is most probably explained by miss-predictions of the signal peptide cleavage site, while in some cases the phenylalanine within the ''ASAFAP''-motif is replaced by tryptophan, phenylalanine or leucine, which are shown to be functional in this study. Bioinformatic approaches to determine plastid proteomes can be impeded by such miss-predictions. Our results will facilitate future bioinformatic analysis of plastid proteomes on a genomic level, since the presence of the ''ASAFAP''-motifs in proximity to a predicted cleavage site can be helpful to test large numbers of proteins for the presence of plastid targeting signals in diatoms and related algae.

The ''ASAFAP''-motif is very conserved in *P. tricornutum* and similar motifs are found in other groups of algae with secondary plastids like dinoflagellates and cryptophytes. Plastid preproteins in dinoflagellates possess a conserved ''FVAP'' motif (Patron et al. 2005), while in cryptophytes ''AXAF'' is found (Gould et al. 2006a). Bipartite presequences containing an ''ASAFAP''-motif apparently are functional across the species border, as several heterologous plastid targeting presequences from

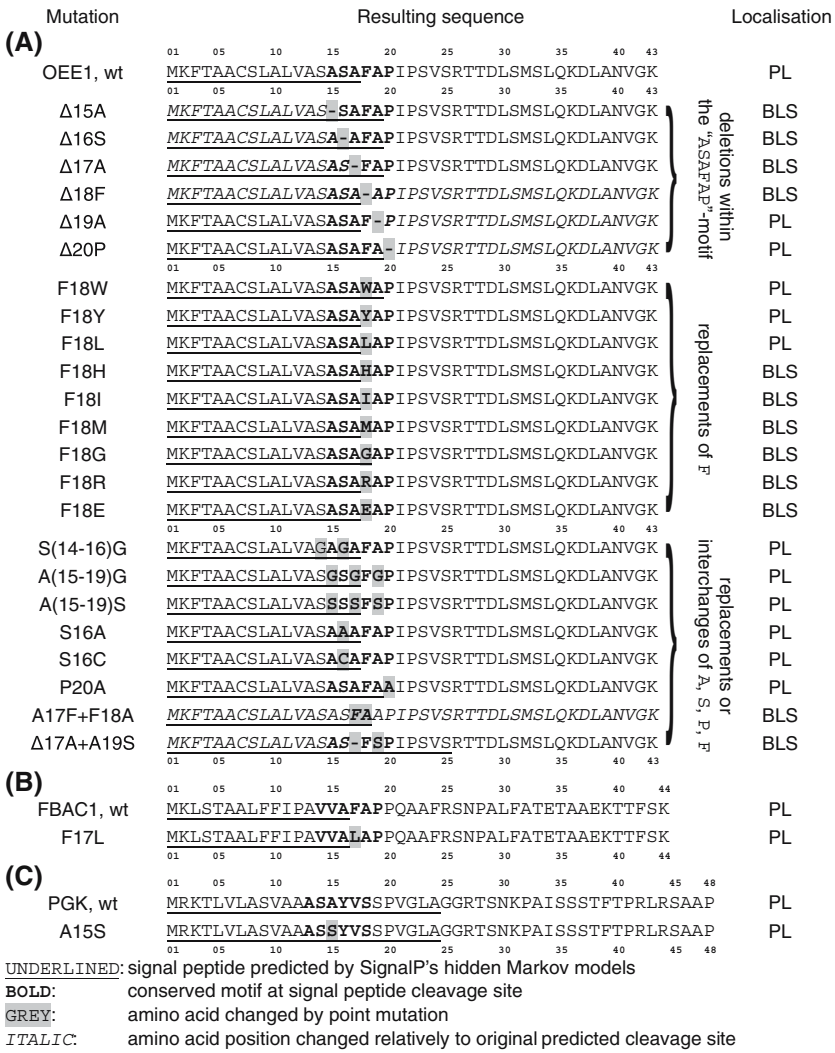


**Fig. 4** Modified presequences generated in this study and localisation of the fusion proteins after expression in *Phaeodactylum tricornutum*. wt: wild type, PL: plastid, BLS: ''blob''-like structure. **(A)** Wild type and modified presequences of the OEE1 (oxygen evolving enhancer protein 1) from *P. tricornutum*. **(B)** Wild type and modified presequence of the PtFBAC1 (fructose-1,6-bisphosphate aldolase) from *P. tricornutum*. **(C)** Wild type and modified presequence of the PGK (phosphoglycerate kinase) from *Guillardia theta*

other diatoms, cryptophytes and dinoflagellates fused to GFP lead to plastid import in *P. tricornutum* (Gould et al. 2006a; Kilian and Kroth 2005; Kroth et al. 2005; Lang 2000). There is even good evidence that the presence of a conserved phenylalanine possibly is not restricted to algal groups with secondary plastids, recently it has been shown that red algae and glaucophytes—both possessing primary plastids—have a consensus sequence with phenylalanine at position three or four at the N-termini of their plastid targeting transit peptides (Steiner and Löffelhardt 2005). At least in glaucopyhtes this phenylalanine has been shown to be crucial in *in vitro* import experiments and may eventually be replaced only by tyrosine (Steiner et al. 2005). The ''ASAFAP''-motif might therefore be a specialised form of a more loosely conserved but widely spread presequence-motif of ''non-green'' algal groups.

The mode of protein translocation into secondary plastids of diatoms is still under debate (Kilian and Kroth 2003). A ''vesicular shuttle model'' (Gibbs 1979) and a ''translocator model'' (Cavalier-Smith 1999, 2003; McFadden 1999) are discussed. Common to both models is the postulation of cotranslational transport across the outermost CER membrane and translocation over the innermost envelope membrane by a Tic related translocon. The models differ in the way they explain the passage of the proteins across the second and the third membrane (counting from outside). The ''vesicular shuttle model'' postulates vesicular transport across the periplastidic space between these membranes, because of vesicles that have been found in the periplastidic space by electron microscopy (Gibbs 1979). The ''translocator model'' proposes that preproteins enter the periplastidic space by translocators or pores and then are imported into the plastid across the residual two membranes via a Tic/Toc system similarly to land plant plastids. A translocator derived from a duplicated Toc or Tic system or an unspecific pore have been suggested to be involved in protein translocation from the CER to the periplastidic space (Cavalier-Smith 1999;

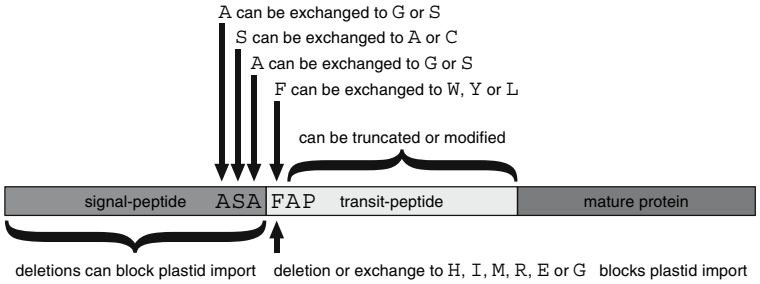


**Fig. 5** Scheme of the presequence structure of diatom plastid preproteins. The phenylalanine at the first position of the transit peptide-like domain can only be replaced by the aromatic amino acids tryptophan and tyrosine or by the large and hydrophobic leucine. Amino acid exchanges at other positions do not affect plastid import. The transit peptide-like domain can be truncated to a large extent, while deletions in the signal peptide can cause a block of plastid import

Kroth and Strotmann 1999). Independent of which model is correct, it is likely that the ''ASAFAP''-motif and the transit peptide-like domain act as signals for actively sorting plastid proteins out of the ER/CER and for further transport into the plastids. It has been shown that it is possible to use a signal peptide fused to ''FATTP'' to target GFP into the plastids, while a signal peptide fused to ''FA'' alone fails to do so and leads to the BLS phenotype (Kilian and Kroth 2005). The fact that a phenylalanine alone or a transit peptide-like domain without phenylalanine led to the BLS phenotype when fused to GFP and expressed in *P. tricornutum* illustrates that both elements are necessary.

The high conservation of phenylalanine and its crucial role for the import reaction indicates that an intracellular receptor/transport system might be involved that recognises a phenylalanine at the N-terminus of the cargo protein. A component derived from the bacterial outer membrane protein Omp85 was proposed to act as phenylalanine specific receptor and membrane channel (Steiner and Löffelhardt 2005). A specific interaction of aromatic amino acid residues within the protein cargo with transport components is also known from protein sorting into caveolae, plasma membrane structures formed in the process of endocytosis (Couet et al. 1997) and from targeting from the trans Golgi network to the vacuole (Bryant and Stevens 1998), but in these cases the interacting aromatic residues are not found at the very N-termini of the cargo proteins. The sequence $F(X)_6LL$ (with X being any residue, and L being either leucine or isoleucine) in the membrane-proximal carboxyl termini of many G protein-coupled receptors mediates receptor protein transport from the ER to the cell surface. However, the precise molecular mechanism by which the $F(X)_6LL$ motif regulates G protein-coupled receptor protein export from the ER is unknown (Duvernay et al. 2004). Since these vague similarities between cargo protein motifs point to unknown mechanisms, conclusions from the ''ASAFAP''-motif on the import mechanism for diatom or cryptophyte plastid proteins remain speculative.

Secretory transport might have been the first protein import system into early primary plastids, which first may have developed the Tic and then the Toc complex (Kilian and Kroth 2003). The ''ASAFAP''-motif may therefore even be a relic of a former import system being present in the ancestor of all plastids before the transit peptide system was developed. Subsequently the strict phenylalanine dependence was overcome in green algae, while red algae and glaucophytes retained the phenylalanine dependent type of import receptor. More evidence for the presence of parallel import pathways in the same organisms comes from the recent discovery that there is a second pathway for chloroplast import in green plastids via the secretory pathway (Villarejo et al. 2005), which might also exist in red algae and which may possibly have been adapted as the main pathway of protein import into secondary red plastids instead of the Tic/Toc-dependent system.

Analyses of the genomes of the diatoms *Thalassiosira pseudonana* and *P. tricornutum* revealed the presence of putative components of the Tic apparatus, but no subunits of the Toc apparatus were identified (Armbrust et al. 2004; McFadden and van Doren 2004; Gruber and Kroth, unpublished). However, we were also not able to detect proteins that might be involved in vesicular transport within the periplastidic space in diatoms up to now, although they should be easily distinguishable from their cytosolic counterparts by the presence of a signal peptide. So from the genome sequence analyses neither the ''vesicular shuttle model'' nor the ''translocator model'' are favoured. It can therefore also be speculated that new or modified systems account for the protein transport over the second and the third membrane and neither a Toc translocon nor vesicular transport are involved. A mitochondrial translocon component, Tim23, was therefore also proposed as possible origin of a translocon involved in the protein translocation out of the ER lumen (Bodył

2004). Furthermore, genes for components of the ER-associated degradation machinery (ERAD) were recently found on the nucleomorph genome of *G. theta*. Respective genes are also duplicated in the genomes of *P. tricornutum* and *T. pseudonana*. An altered ERAD-related machinery involved in the regular transport of properly folded proteins out of the ER and into the periplastidic compartment was therefore suggested (Sommer et al. 2007). Meanwhile considerable knowledge about the presequence structure of nucleus encoded plastid targeted proteins from diatoms, cryptophytes and dinoflagellates was gained (Apt et al. 2002; Gould et al. 2006a; Kilian and Kroth 2005; Nassoury et al. 2003; Patron et al. 2005, this study), remarkably, the detailed import process of proteins targeted to the plastids via the ER remains largely unknown.

**Acknowledgements** We thank D. Ballert for help with the transformation and cultivation of *Phaeodactylum tricornutum*. This study was supported by the University of Konstanz and grants of the Deutsche Forschungsgemeinschaft (Project KR 1661/3) and the European community (MARGENES, project QLRT-2001-01226) to PGK.

## References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–402

Apt KE, Kroth-Pancic PG, Grossman AR (1996) Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol Gen Genet 252:572–579

Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O et al (2002) *In vivo* characterization of diatom multipartite plastid targeting signals. J Cell Sci 115:4061–4069

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D et al (2004) The genome of the Diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795

Bhaya D, Grossman A (1991) Targeting proteins to diatom plastids involves transport through an endoplasmic reticulum. Mol Gen Genet 229:400–404

Bodył A (2004) Evolutionary origin of a preprotein translocase in the periplastid membrane of complex plastids: a hypothesis. Plant Biol 6:513–518

Bryant N, Stevens T (1998) Vacuole biogenesis in *Saccharomyces cerevisiae*: protein transport pathways to the yeast vacuole. Microbiol Mol Biol Rev 62:230–247

Cavalier-Smith T (1999) Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate and sporozoan plastid origins and the eukaryotic family tree. J Eukary Microbiol 46:347–366

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

Cavalier-Smith T (2003) Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Philos Trans R Soc Lond B Biol Sci 358:109–134

Chaal BK, Green BR (2005) Protein import pathways in 'complex' chloroplasts derived from secondary endosymbiosis involving a red algal ancestor. Plant Mol Biol 57:333–342

Couet J, Li S, Okamoto T, Ikezu T, Lisanti M (1997) Identification of peptide and protein ligands for the caveolinscaffolding domain. Implications for the interaction of caveolin with caveolae-associated proteins. J Biol Chem 272:6525–6533

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. Am Nat 154:S164–S177

Douglas SE, Penny SL (1999) The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol V48:236–244

Duvernay MT, Zhou F, Wu G (2004) A conserved motif for the transport of G protein-coupled receptors from the endoplasmic reticulum to the cell surface. J Biol Chem 279:30741–30750

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978–984

Gibbs SP (1979) The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. J Cell Sci 35:253–266

Gibbs SP (1981) The chloroplast endoplasmic reticulum: structure, function and evolutionary significance. Int Rev Cytol 72:49–99

Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG et al (2006a) Protein targeting into the complex plastid of cryptophytes. J Mol Evol V62:674–681

Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ et al (2006b) Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. Mol Biol Evol 23:2413–2422

Harper JT, Waanders E, Keeling PJ (2005) On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J System Evol Microbiol 55:487–496

Heazlewood JL, Tonti-Filippini J, Verboom RE, Millar AH (2005) Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis. Plant Physiol 139:598–609

Ishida K (2005) Protein targeting into plastids: a key to understanding the symbiogenetic acquisitions of plastids. J Plant Res 118:237–245

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

Keeling PJ (2004) Diversity and evolutionary history of plastids and their hosts. Am J Bot 91:1481–1493

Kilian O, Kroth PG (2003) Evolution of protein targeting into ''complex'' plastids: the ''secretory transport hypothesis''. Plant Biol 5:350–358

Kilian O, Kroth PG (2004) Presequence acquisition during secondary endocytobiosis and the possible role of introns. J Mol Evol 58:712–721

Kilian O, Kroth PG (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J 41:175–183

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kroth PG (2002) Protein transport into secondary plastids and the evolution of primary and secondary plastids. Int Rev Cytol 221:191–255

Kroth PG (2007) Genetic transformation; a tool to study protein targeting in diatoms, chap. 17. In: Methods in molecular biology, 2nd edn., Totowa, NJ, USA: Humana Press

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

Kroth PG, Schroers Y, Kilian O (2005) The peculiar distribution of class I and class II aldolases in diatoms and in red algae. Curr Genet 48:389–400

Lang M (2000) Untersuchungen zum Transport kernkodierter Plastiden-Proteine in Kieselalgen. Ph.D. thesis, Heinrich-Heine-Universität Düsseldorf, URL http://diss.ub.uni-duesseldorf.de/home/etexte/diss/file?dissid=37

Lang M, Apt KE, Kroth PG (1998) Protein transport into ''complex'' diatom plastids utilizes two different targeting signals. J Biol Chem 273:30973–30978

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY et al (2005) CDD: a conserved domain database for protein classification. Nucl Acids Res 33:D192–D196

Martin W, Herrmann RG (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 et al (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165

McFadden GI (1999) Plastids and protein targeting. J Eukaryot Microbiol 46:339–346

McFadden GI (2001) Primary and secondary endosymbiosis and the origin of plastids. J Phycol 37:1–9

McFadden GI, van Dooren GG (2004) Evolution: red algal genome affirms a common origin of all plastids. Curr Biol 14:R514–R516

Millar AH, Whelan J, Small I (2006) Recent surprises in protein targeting to mitochondria and plastids. Curr Opin Plant Biol 9:610–615

Montsant A, Jabbari K, Maheswari U, Bowler C (2005) Comparative genomics of the pennate diatom *Phaeodactylum tricornutum*. Plant Physiol 137:500–513

Moreira D, Le Guyader H, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. Nature 405:69–72

Nassoury N, Cappadocia M, Morse D (2003) Plastid ultrastructure defines the protein import pathway in dinoflagellates. J Cell Sci 116:2867–2874

Nielsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol 6:122–130

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997a) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997b) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst 8:581–599

Oudot-Le Secq MP, Grimwood J, Shapiro H, Armbrust EV, Bowler C, et al. (2007) Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. Mol Genet Genom 277(4):427–439. PMID: 17252281

Pancic PG, 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

Patron NJ, Waller RF, Archibald JM, Keeling PJ (2005) Complex protein targeting to dinoflagellate plastids. J Mol Biol 348:1015–1024

Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G et al (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 15:1325–1330

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097–6100

Soll J, Schleiff E (2004) Protein import into chloroplasts. Nat Rev Mol Cell Biol 5:198–208

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM et al (2007) Der1-mediated pre-protein import into the periplastid compartment of chromalveolates? Mol Biol Evol 24(4):918–928. PMID: 17244602

Starr RC, Zeikus JA (1993) UTEX: the culture collection of algae at the University of Texas at Austin, 1993 list of cultures. J Phycol 29:1–106

Steiner JM, Löffelhardt W (2005) Protein translocation into and within cyanelles. Mol Membr Biol 22:123–132

Steiner JM, Yusa F, Pompe JA, Löffelhardt W (2005) Homologous protein import machineries in chloroplasts and cyanelles. Plant J 44:646–652

Timmis JN, Ayliffe MA, Huang CY, Martin W (2005) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135

Villarejo A, Buren S, Larsson S, Dejardin A, Monne M et al (2005) Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat Cell Biol 7:1224–1231

von Heijne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133:17–21

Waller RF, McFadden GI (2004) The Apicoplast, chap. 11. Caister Academic Press, Wymondham, UK, pp 291–338

Wastl J, Maier UG (2000) Transport of Proteins into cryptomonads complex plastids. J Biol Chem 275:23194–23198

Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE (2000) Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. J Phycol 36:379–386