In Silico and In Vivo Investigations of Proteins of a Minimized Eukaryotic Cytoplasm

Daniel Moog[1,2], Simone Stork[1,2], Stefan Zauner[1,2], and Uwe-G. Maier*[,1,2]

[1]LOEWE-Zentrum für Synthetische Mikrobiologie (SynMikro), Hans-Meerwein-Strasse, Marburg, Germany

[2]Laboratory for Cell Biology, Philipps University Marburg, Karl-von-Frisch Strasse 8, Marburg, Germany

*Corresponding author: E-mail: maier@staff.uni-marburg.de.

Accepted: 29 March 2011

Abstract

Algae with secondary plastids such as diatoms maintain two different eukaryotic cytoplasms. One of them, the so-called periplastidal compartment (PPC), is the naturally minimized cytoplasm of a eukaryotic endosymbiont. In order to investigate the protein composition of the PPC of diatoms, we applied knowledge of the targeting signals of PPC-directed proteins in searches of the genome data for proteins acting in the PPC and proved their in vivo localization via expressing green fluorescent protein (GFP) fusions. Our investigation increased the knowledge of the protein content of the PPC approximately 3-fold and thereby indicated that this narrow compartment was functionally reduced to some important cellular functions with nearly no housekeeping biochemical pathways.

Key words: secondary endosymbiosis, periplastidal compartment (PPC), plastid protein import, bipartite targeting signal (BTS), *Phaeodactylum tricornutum*, diatom.

Introduction

The cytoplasm is an essential compartment with many functions. However, algae with plastids of secondary origin which are surrounded by four membranes harbor two evolutionarily different cytoplasms per individual cell (Cavalier-Smith 1999). The additional cytoplasm originated from the integration of a phototrophic eukaryotic cell into another eukaryotic one. Here, successive reduction of the endosymbiont led to a complex plastid surrounded by either three or four membranes, as it is found in many organisms of ecological or medical interest, such as cryptophytes, chlorarachniophytes, heterokonts, haptophytes, euglenophytes, peridinin-containing dinoflagellates, and apicomplexa (Hempel et al. 2007; Bolte et al. 2009).

In organisms with complex plastids surrounded by four membranes, the outermost membrane might trace back to a phagotrophic membrane, which is in several phyla fused with the endoplasmic reticulum (ER) of the host. The second outermost membrane (periplastidal membrane [PPM]) resembles the former plasma membrane of the eukaryotic endosymbiont, and both innermost membranes are homologous to the plastid envelope of archaeplastida (Cavalier-Smith 2003). Thus, the space between the second and third outermost membrane represents the cytoplasm of the eukaryotic endosymbiont (fig. 1*B*). In cryptophytes and chlorarachniophytes, this remnant compartment, called periplastidal compartment (PPC), harbors a pigmy cell nucleus, the nucleomorph, which was shown to be the remnant nucleus of the respective eukaryotic endosymbiont (Maier et al. 2000; Douglas et al. 2001; Gilson et al. 2006; Lane et al. 2007). However, most of secondarily evolved organisms show no obvious compartmentalization in the PPC. Thus, nature provided an interesting example for a step-by-step reduction of a cytoplasmic compartment, namely a reduction series from a cytoplasm in a free-living eukaryote to a reduced, but genetically active PPC in cryptophytes and chlorarachniophytes, or even further to a pigmy cytoplasm devoid of a nucleomorph and therefore without genetic activities in heterokonts, haptophytes, and apicomplexa.

In previous work, we and others have characterized the targeting signals, which are important for directing nucleus-encoded proteins into the PPC of cryptophytes and diatoms (Gould et al. 2006a; Gruber et al. 2007; Sommer et al. 2007). It was shown that PPC-imported proteins are equipped with an N-terminal bipartite targeting sequence (BTS), composed of a signal peptide (SP) followed by a transit peptide-like sequence (TP), in which the first amino acid (aa) is not aromatic or a leucine (Kilian and Kroth 2005; Gould et al. 2006a,

© The Author(s) 2011. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution*.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

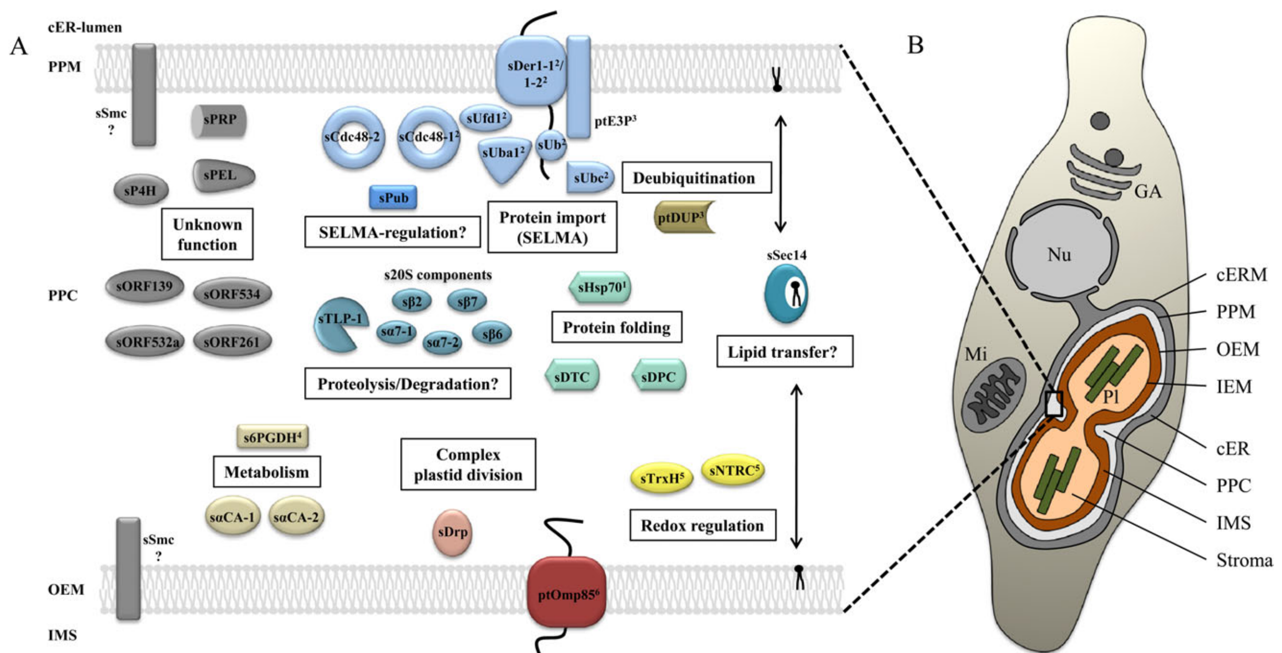

Fig. 1.—The PPC and cell structure/plastid compartmentalization of the diatom *P. tricornutum*. (*A*) Postulated functions of PPC-localized proteins are indicated in black boxes. s, symbiontic (PPC-localized); sDer1-1/1-2, degradation at the ER; sUfd1, ubiquitin fusion degradation; sUba1, ubiquitin activating E1; sUbc, ubiquitin conjugating E2; ptE3P, *P. tricornutum* ubiquitin ligase E3 of the PPC; Ub, ubiquitin; ptDUP, *P. tricornutum* deubiquitinating enzyme of the PPC; sCdc48-1/2, cell division cycle protein; sPUB, PUB and thioredoxin domain containing protein; sHsp70, heat shock protein; sDTC, DnaJ and TPR domain-containing protein; sDPC, DnaJ and PDI domain-containing protein; sSec14, putative lipid transfer protein sTrxH, thioredoxin; sNTRC, NADPH depending thioredoxin reductase containing N-terminal thioredoxin domain; sDrp, dynamin-related protein; sα7/sβ2/6/7, proteasomal 20S components of the alpha and beta type; sTLP-1, trypsin-like serine protease; sSMC, structural maintenance of the chromosome-like protein; sPRP, pentapeptide repeats containing protein; sPEL, pectin esterase domain-containing protein; sP4H, prolyl-4-hydroxylase; 6PGDH, 6-phosphoglucono-lactone dehydrogenase; sαCA-1/2, alpha carbonic anhydrase; sORF139/261/532a/534, open reading frame (*Guillardia theta* nucleomorph-encoded ORF homolog); ptOmp85, outer membrane protein. Superscript numbers indicate proteins localized in previous studies: 1, (Gould et al. 2006a); 2, (Sommer et al. 2007); 3, (Hempel et al. 2010); 4, (Gruber et al. 2009); 5, (Weber et al. 2009); 6, (Bullmann et al. 2010). (*B*) *Phaeodactylum tricornutum* possesses an aliform-shaped secondary plastid surrounded by four membranes. The outermost membrane is in continuum with the host rough endoplasmic reticulum (rER). The space between the two inner- and outermost membrane pairs (PPC) represents the former red algal cytoplasm of the endosymbiont. cERM, chloroplast ER membrane; cER, chloroplast ER; PPM, periplastidal membrane; PPC, periplastidal compartment; OEM, plastid outer envelope membrane; IMS, plastid intermembrane space; IEM, plastid inner envelope membrane; Pl, plastid; Nu, nucleus; Mi, mitochondrion; GA, Golgi apparatus.

2006b; Gruber et al. 2007). In order to investigate the biochemical and cell biological capacities of the PPC of the model organism *Phaeodactylum tricornutum*, we screened the genomic data base of the diatom for candidates possessing a PPC-specific BTS and proved positives by in vivo targeting experiments. By combining the available data of PPC-localized proteins with the here generated new data set, we present the first compilation of present and absent functions to a minimized eukaryotic cytoplasm.

Materials and Methods

Bioinformatical Analysis

To identify putative PPC-localized proteins, we searched for components involved in cytosolic processes and analyzed them for the presence of an N-terminal BTS. Protein sequences were either retrieved by direct search from the KOG classification of the *P. tricornutum* data base (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html), the National Center for Biotechnology Information (NCBI) protein data base (http://www.ncbi.nlm.nih.gov/protein), or the *P. tricornutum* data base was screened by Blast search for specific proteins using, unless otherwise noted, protein sequences from *Saccharomyces cerevisiae* (http://www.yeastgenome.org/), *Arabidopsis thaliana* (http://www.arabidopsis.org/), or *Cyanidioschyzon merolae* (http://merolae.biol.s.u-tokyo.ac.jp/) as queries. Proteins were classified based on retrieved NCBI Blast hits and conserved domains identified by the NCBI conserved domain search ((http://www.ncbi.nlm.nih.gov/cdd). Predicted gene models were examined based on available expressed

sequence tag (EST) data to determine the correct protein sequence. To identify putative PPC-localized proteins, three criteria were considered.

First, sequences were screened for the presence of a BTS, starting with the SP prediction with SignalP (http://www.cbs.dtu.dk/services/SignalP/) and TargetP (http://www.cbs.dtu.dk/services/TargetP/), distinguishing between predicted cytoplasmic, secretory, and mitochondrial localizations. Proteins containing a predicted SP were then analyzed for a putative transit peptide-like sequence using TargetP (http://www.cbs.dtu.dk/services/TargetP/) and ChloroP (http://www.cbs.dtu.dk/services/ChloroP/). Due to the weak prediction performance of TargetP and ChloroP for transit peptide-like sequences of organism that are of secondary endosymbiotic origin, putative candidates were checked for N-terminal extensions by determining the conserved regions of the mature proteins by NCBI Blast search. Several amino acids separating the SP from the conserved region were defined as sufficient for a putative second part of the BTS. Regarding the second criterion, essential proteins had to be present in at least two copies in the genome of *P. tricornutum* in order to assure the specific function in the host cytosol. Third, proteins not known to contain a SP—like exclusively cytosolic functions—but having one in *P. tricornutum* were considered as putative PPC proteins irrespective of the lengths of a transit peptide-like sequence.

Plasmid Construction and Transfection of *P. tricornutum*

Those genes, which were predicted to encode PPC proteins, were cloned and transfected into the diatom *P. tricornutum*. Here, either the BTS or the full-length coding sequence was fused to *egfp*, depending on whether the end of the putative BTS could be clearly differentiated from the mature protein part indicated by Blast analysis (conservation of the protein at the N-terminus). With some exceptions (indicated in the supporting information S2, Supplementary Material online), the *P. tricornutum* sequences were amplified from gDNA or cDNA if the predicted gene model wasn't confirmed with EST data, using standard polymerase chain reaction (PCR) conditions. Genomic sequences can be retrieved from the *P. tricornutum* database (PhatrDB v2.0). For further information about the protein sequences used for transfection and primer sequences used for PCR, see supporting information S2 and S3 (Supplementary Material online). For eGFP localization studies, either the BTS or the full-length coding sequences (as explained above) were cloned in front of *egfp* into the pPha-T1 vector and biolistically transfected into *P. tricornutum* cells as described previously (Zaslavskaia et al. 2000; Sommer et al. 2007).

Confocal Microscopy

All *P. tricornutum* transformants were analyzed with a confocal laser scanning microscope Leica TCS SP2 using a HCX PL APO 40×/1.25 − 0.75 Oil CS objective. Fluorescence of eGFP and chlorophyll was exited with an Argon laser at 488 nm and detected with two photomultiplier tubes at a bandwidth of 500–520 nm and 625–720 nm for eGFP and chlorophyll fluorescence, respectively.

Results and Discussion

The PPC is a naturally minimized eukaryotic cytoplasm found in organisms with plastids surrounded by four membranes (Gould et al. 2008). Regardless of the different phylogenetic origin of the organisms from which the PPC originated (green algal endosymbiont in chlorarachniophytes, or red algal endosymbiont in heterokontophytes, haptophytes, cryptophytes, and apicomplexa) (Archibald 2009), two fundamentally different types exist. On one hand, the PPC of cryptophytes and chlorarachniophytes has the capacity for protein biosynthesis as shown by the presence of a transcriptionally active nucleomorph and 80S ribosomes. On the other hand, the PPC of all other organism with secondary plastids lack a genetic apparatus (Keeling 2009). Here, we have investigated the PPC of a diatom (fig. 1*B*) to learn more about biochemical and cell biological capacities of a naturally reduced cytoplasm.

Data Mining

The PPC in diatoms is not directly/biochemically accessible for proteome analyses so far. Thus, we used a combined in silico/in vivo-localization approach to determine proteins localized to the PPC.

As a first step, we extracted from the filtered model (best model, proteins–chromosomes) data set of the *P. tricornutum* data base (http://genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html), those protein models exceeding a 50% cutoff in the hidden markov model SP prediction of the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/). Starting with 10,025 protein models, this approach led to a list of 2,260 entries all showing a putative SP. After excluding all putative plastid localized models characterized by the presence of either an aromatic amino acid or leucine at the first position of the putative TPs (970 models), a list of 1,290 protein models remained.

The computer-based automatic identification of a transit peptide-like sequence (after removal of the predicted SP) turned out to be difficult as bioinformatic tools like TargetP and ChloroP have not been trained for the TP prediction of secondarily evolved organisms. Thus, the automatic discrimination between secretory proteins (without TP) and those having a PPC-specific BTS (with TP) has not turned out effectual. Additionally tests on randomly chosen predicted

gene models often resulted in a problematic detection of the genuine N-terminus of the models especially of such proteins with unknown function and/or poor conservation. For these reasons, we changed our strategy and focused on the direct identification of proteins known to be involved in specific cellular/cytosolic processes. In particular, we searched for proteins involved in vesicular trafficking as well as on soluble factors involved in lipid biosynthesis, lipid transfer, glycerophospholipid-/glycerolipid-metabolism, CO_2 concentration regulation, chaperones, and prolyl-isomerases. In addition, we inspected entries for components of the cytoskeleton and for proteins acting in plastid division as well as in protein transport. We also searched for kinases and phosphatases and components of the glutathione system. Last but not least, we were interested in cryptochromes and homologues to components encoded by the nucleomorph of a cryptophyte. We excluded factors involved in carbohydrate metabolism, as this was already investigated by another group for the diatom (Kroth et al. 2008), as well as membrane proteins in most cases.

By this strategy (see experimental procedures for details), we identified 467 genes and their encoded products of *P. tricornutum* (supplementary table S1, Supplementary Material online). All these proteins were manually checked for the presence of a BTS with the specificity for PPC-localization (Gould et al. 2006b, 2008; Gruber et al. 2007). A putative PPC-localization was predicted for 50 entries. Forty of them were analyzed by expressing either the BTS or the complete gene (depending on the fact whether the length of the putative BTS could be clearly identified by N-terminal conservation of the mature protein) as eGFP fusion protein in *P. tricornutum*. It was shown earlier that a punctated eGFP signal adjacent to the chlorophyll autofluorescence of the plastid, originally termed as "blob-like structure" (Kilian and Kroth 2005), indicates a PPC-localization. Such a PPC-specific eGFP signal was obtained in 22 cases (fig. 2). Ten further fusion proteins entered the secretory pathway, whereas six constructs showed a mitochondrial localization. In one case, a plastidal signal could be obtained as well as one construct with a cytosolic eGFP localization. Taken together, about three-quarter of the predicted proteins enter the secretory pathway but only approximately 55% of the predicted PPC proteins are actually localized in the PPC, indicating the limits of available bioinformatical tools.

Equally important as PPC positives are functions not detectable in the minimized cytoplasm. From the cellular functions searched for, we identified no PPC-specific components involved in vesicular trafficking (Rabs, SNAREs, COPI and COPII, Clathrin, Caveolin, ESCRT, GEFs, and GAPs), cytoskeleton, regulatory components such as kinases and phosphatases, cryptochromes, and enzymes catalyzing lipid biosynthesis (see below). A negative result might be caused for an individual gene by an incorrect predicted gene model or by the divergence of the PPC-specific components. However, for complex cellular functions, such as vesicular trafficking or proteins, which are conserved in eukaryotes such as actin, a negative result, that is, no entry detected, support but not finally prove the absence of a PPC-specific expressed version.

Even more important, false positives might be mainly caused by limits of the prediction programs. As long as there is no prediction tool with high confidence available for organism containing organelles of secondary origin, in vivo localizations are absolutely necessary. This is generally not due to the prediction of the SPs, instead, the low quality for prediction is mainly based on the not conserved and poorly characterized transit peptide-like sequences.

Protein Import into and Export Out of the PPC

The PPC has to be crossed by hundreds of nucleus-encoded plastid proteins. According to recent findings, protein transport across the second and third outermost membranes is proposed to be mediated by two protein translocons (Sommer et al. 2007; Bullmann et al. 2010). These are a symbiont-specific ERAD-like machinery (SELMA), a modified ERAD system in the second outermost membrane (PPM) (fig. 1*A*), which is composed of membrane proteins (sDer1-1, sDer1-2, ptE3P) and accompanied by soluble PPC proteins (sCdc48, sUfd1, sUba1, sUbc, sUb, ptDUP) to be functional (Agrawal et al. 2009; Hempel et al. 2009; Kalanon et al. 2009; Spork et al. 2009). In the third outermost membrane (OEM), an Omp85 protein has been identified (Bullmann et al. 2010) (fig. 1*A*). An additional SELMA factor might be a second version of Cdc48, sCdc48-2, which we identified here (fig. 2). It shares a high sequence similarity to sCdc48-1 and the cytosolic host version that functions in ERAD. It is known that Cdc48 forms homo-oligomers in the genuine ERAD (Aker et al. 2007); the presence of a second copy in the diatoms PPC offers the possibility for the formation of hetero-oligomers of both proteins (sCdc48-1/2) in SELMA. Another possibility might be that one of the PPC-located sCdc48 has adopted some other functions aside (or connected to) the SELMA system (see below).

With sPUB, we identified a further PPC-located protein in *P. tricornutum* (fig. 2) which, in respect of its PUB-domain, might be involved in SELMA functions/transport (fig. 1*A*). The PUB-domain usually is found in eukaryotic proteins closely linked to the ubiquitin-proteasome system (e.g., PNGases, Uba, Ubx) (Suzuki et al. 2001; Allen et al. 2006; Madsen et al. 2009) and is known to be an interaction module for Cdc48. Because of its additional thioredoxin domain, a redox regulatory function of sPUB for SELMA via sCdc48-1/2 and therefore for protein import into the PPC might be presumed. sPUB might act in concert with a periplastidal thioredoxin (sTrxH2) and thioredoxin reductase (sNTRC), which were recently identified in a PPC version (fig. 1*A*) by the group of Peter Kroth (Weber et al. 2009).

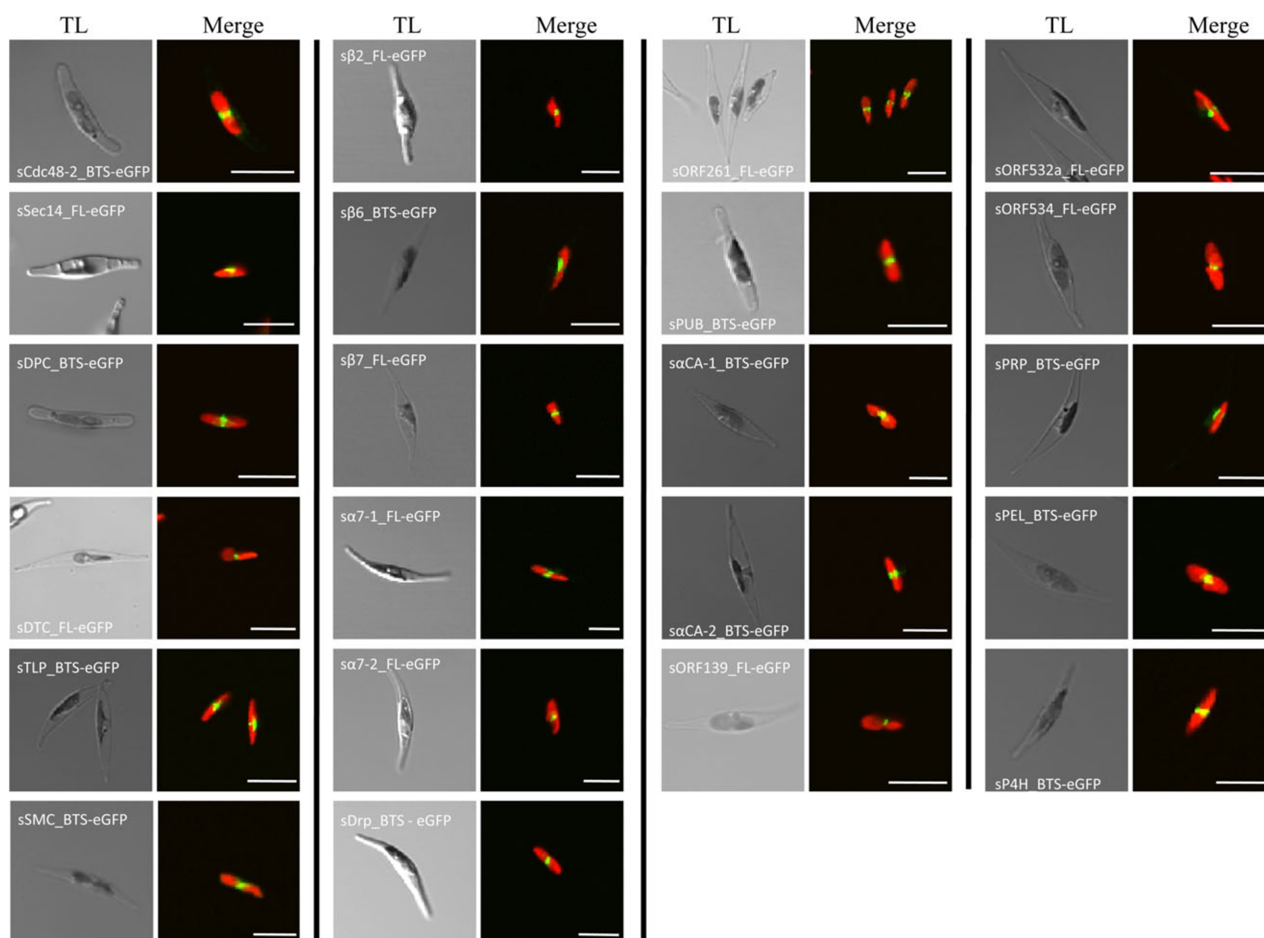

Fig. 2.—In vivo localization studies of *P. tricornutum* BTS/FL sequences fused to eGFP. Homologous overexpression of BTS- or full-length (FL)-GFP fusion proteins in all cases led to a characteristic 'blob-like' GFP-fluorescence pattern, known to correspond to a typical PPC-localization. The blob-like structure is due to a median constriction of the two innermost membranes of the plastid (OEM/IEM), which leads to a widening of the PPC. TL, transmitted light; Merge, overlay of plastid autofluorescence (red) and GFP fluorescence (green). The scale bar represents 10 μm. For further information about the proteins, see text and the legend of figure 1*A*.

Because we failed to identify a PPC-directed glutathione system (supplementary table S1, Supplementary Material online), the thioredoxin system might be sufficient for maintaining the redox state of the PPC.

Vesicular Trafficking Was Not Detected in the PPC of the Diatom

As mentioned earlier (Hempel et al. 2009), we cannot definitively exclude the possibility that nucleus-encoded proteins destined either for the PPC or the plastid stroma use different routes for crossing the second outermost membrane. This might be indicated by vesicular structures between the plastid surrounding membranes, which were detected in electron microscopical studies (Gibbs 1979). As vesicular protein transport through the PPC might be an alternative route to the plastid (Gibbs 1979), we screened the database for factors involved in vesicle generation and fusion (COPI, COPII, Clathrin, Caveolin, ESCRT, SNAREs, Rabs, GEFs, and GAPs). Whereas we were able to detect host copies in most cases (see supplementary table S1, Supplementary Material online), candidates for PPC-located members of these protein complexes were not identified or shown to be wrongly predicted as PPC proteins. In addition, we did not detect any PPC-located components of actin or tubulin. These negative results led us to speculate that vesicles with a protein composition known to be important for vesicular transport are not present in the PPC. Consequently, the vesicular structures observed in heterokontophytes might have other functions than vesicle-mediated protein transport.

In the exterior layer of the outer envelope membrane of primary plastids, eukaryotic phospholipids, synthesized at the ER, replaced the former cyanobacterial lipopolysaccharide during primary endosymbiosis (Cavalier-Smith 2000). Because the innermost membrane pair of secondary plastids

is homologous to the primary plastid envelope, they should depend on eukaryotic membrane lipids as well. As typical vesicles might not be present in the PPC of the diatom, lipid exchange between host and symbiont membranes has to be organized by a different mechanism. We found no indications for lipid biosynthesis in the PPC (supplementary table S1, Supplementary Material online). Therefore, we screened for lipid transfer proteins, which succeeded in the identification of a homolog to the phosphatidylcholine/phosphatidylinositol transporting protein Sec14p from *S. cerevisiae* (Mousley et al. 2007), which is PPC-located in vivo (fig. 2).

Protein Folding

Recently, we have shown that the putative translocon in the third outermost membrane, ptOmp85, can be passed by unfolded proteins only (Bullmann et al. 2010). Thus, keeping proteins in a transport-competent conformation might be an important issue in the PPC. We already reported that a copy of Hsp70 is PPC localized (Gould et al. 2006b; Sommer et al. 2007). Additional hits for factors involved in protein folding were obtained for sDTC, a probable Hsp70 cochaperone containing an Hsp40-like DnaJ-domain and several tetratricopeptide repeats, which are known to mediate protein–protein interaction and the assembly of protein complexes, and sDPC, which possesses a DnaJ- and PDI (protein disulphide isomerase)- domain, by bioinformatic search (supplementary table S1, Supplementary Material online). The PPC-localization of both proteins was verified in vivo (fig. 2).

Protein Turnover

The PPC is a reduced cytoplasm and one can expect that a protein degradation/elimination machinery might be present for protein turnover. Protein degradation can be facilitated by different proteases or by the 26S proteasome. We detected several trypsin-like serine proteases (supplementary table S1, Supplementary Material online) with a predicted PPC-targeting signal. In vivo expression of such enzymes might be harmful for the cell. However, expressing only the BTS of one identified trypsin-like protease (sTLP1), a PPC-signal was obtained (fig. 2). In addition, several subunits of a proteasome could be identified with a predicted PPC-targeting signal. Interestingly, only some components of the 20S proteasome core complex (α2, two α7, β2, β3, β6, and β7), but no factors for the 19S cap were detected in a PPC version. The subunits α7-1/2, β2, β6, and β7 were exemplarily shown to be targeted to the PPC (fig. 2). Although the predicted PPC-specific "minimal" proteasome is lacking the regulatory subunits, it might have the capacity to form a cavity, in which proteins can be degraded. This implies that the 20S subunits might be not involved in typical ubiquitin-dependent protein degradation. If so, ubiquitination in the PPC is reserved for protein transport only, and the recently reported deubiquitinating enzyme might be involved in maturation of proteins after ubiquitin-dependent transport across the second outermost membrane (Hempel et al. 2010). In any case, the lack of the 19S regulatory particle raises the question if there are other proteins involved in substrate recognition and unfolding. It was reported that the mammalian AAA-ATPase p97 (homolog of Cdc48) with its cofactors (Ufd1, Npl4) is required for unfolding of some soluble cytoplasmic substrates before degradation (Beskow et al. 2009). Therefore, one of the two symbiontic chaperone-like Cdc48 proteins (see above) might interact with the 20S proteasomal subunits and provide an unfolding activity, possibly sufficient to form a "basic functional proteasome" in the PPC.

Cytoskeleton and Plastid Division

We could not identify genes for actin and for the subunits of tubulin or intermediate filaments in a PPC-directed version. However, a SMC-like protein (structural maintenance of the chromosome-like protein) is present in the PPC, as shown by the in vivo localization of a green fluorescent protein (GFP) fusion with the N-terminal targeting signal (fig. 2). By use of TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/), a transmembrane domain of 20 amino acids at the very end of the C-terminus of sSMC was predicted, indicating that the protein might be anchored to the second (PPM) or third (OEM) outermost membrane of the complex plastid. In the absence of chromosomes in the PPC of diatoms, it might function as a structural element in general.

Division of the complex plastid in *P. tricornutum* might involve cell and organelle division proteins from the former red algal cytoplasm. So far, we could identify a dynamin-related protein (sDrp, fig. 2) belonging to the group of Drp5b/ARC5 proteins, which are involved in division of primary plastids (Gao et al. 2003; Yang et al. 2008), in a PPC-directed version in *P. tricornutum*. Hence, diatoms retained the ARC5 protein in contrast to apicomplexa for which a new family of dynamin-like proteins involved in apicoplast division was recently reported from *Toxoplasma gondii* (van Dooren et al. 2009).

CO_2 Concentration Mechanisms

Enzymes for housekeeping biochemistry, such as glycolysis, oxidative, and reductive pentose phosphate pathway, were investigated by another group (Gruber et al. 2009). However, with 6PGDH, only one PPC-localized protein was verified by in vivo localization studies (Gruber et al. 2009). We additionally searched for carbonic anhydrases (CAs), which in part already have been identified by genome analysis and investigated concerning their subcellular localization in previous studies (Tanaka et al. 2005; Szabo and Colman 2007; Kitao et al. 2008; Kroth et al. 2008). These enzymes catalyze the reversible interconversion of CO_2 and HCO_3^- (Roberts et al. 1997; Raven 2010) and

are crucial components of the inorganic carbon concentrating mechanism (CCM) and CO_2 fixation in diatoms (Tanaka et al. 2005; Kitao et al. 2008). We detected ten CAs in the genome of *P. tricornutum* via bioinformatic analysis (supplementary table S1, Supplementary Material online): Two β-type CAs, which are already known to be plastid localized (Tanaka et al. 2005; Kitao et al. 2008), three γ-CAs and five of the α-type. Several of the α-CAs were predicted as PPC proteins and in vivo localizations of the tested candidates showed that two of them indeed are PPC-localized (fig. 2). Because of the intricate buildup and compartmentalization of the diatoms complex plastid (four envelope membranes), an efficient flux of CO_2 for acquisition and fixation of inorganic carbon by RuBisCO might be much more challenging than in organisms with primary plastids. The presence of CAs in the PPC of *P. tricornutum* might solve this problem on one hand by raising the concentration of CO_2 in the compartment immediately surrounding the plastid and thus in close proximity to RuBisCO (Kroth et al. 2008) and on the other hand by building an efflux barrier for CO_2 from the plastid (Tanaka et al. 2005).

While this manuscript was in revision, Tachibana et al. (2011) published a study in which the cellular localization of CAs of *P. tricornutum* was determined. They reported on nine putative CAs (five α-, two β-, and two γ-CAs) and localized six of them by expressing their estimated N-terminal presequences as GFP fusion proteins in the diatom (Tachibana et al. 2011). In respect to a PPC-localization of CAs, our results are in agreement with their study in the case of one enzyme (sαCA-1). However, the difference in the localization of the second PPC-located CA (sαCA-2) determined by us might be caused by a divergent length of the targeting signal used in each case. Because the N-terminus of the mature CA (sαCA-2) was difficult to determine in silico due to poor conservation, we included the first 169 aa as putative targeting sequence preceding GFP to investigate in vivo localization, whereas in the study of Tachibana et al. (2011), only 46 aa were used. In any case, the study of Tachibana et al. (2011) and ours indicate that CAs are present in the PPC of *P. tricornutum*.

Miscellaneous

Cryptophytes have plastids surrounded by four membranes. However, their PPC is more complex than that of diatoms because the remnant of the nucleus of the secondary endosymbiont, the nucleomorph, is still present in the PPC. The first complete nucleomorph genome sequence was published 2001 (Douglas et al. 2001), indicating a genetically active PPC, which is contrary to that of diatoms. Nevertheless, we used the nucleomorph-encoded genes from the cryptophyte *Guillardia theta* for screening the *P. tricornutum* database. This resulted in the identification of four ORFs, for which homologs are also present in the diatom in a nucleus-encoded but PPC-directed version (fig. 2). Although a functional classification of these ORFs via homology search is inconclusive at the moment, Blast results revealed a general conservation, in species containing a red alga as secondary endosymbiont with the exception of sORF139. Furthermore, we were able to demonstrate a PPC-localization for a pentapeptide repeats containing protein (sPRP), a pectin esterase-like protein (sPEL), and one protein possessing a prolyl-4-hydroxylase domain plus tetratricopeptide repeats (sP4H) (fig. 2), which might be important for PPC maintenance or structure, but the explicit functions in the diatom PPC are unknown yet (fig. 1*A*).

Conclusions

The data presented here provide insights into the composition of the soluble factors of the PPC of the diatom *P. tricornutum* (fig. 1*A*). Our findings on existing and, equally important, missing functions in the PPC of diatoms highlight a naturally minimized compartment with reduced capacities in cellular and biochemical functions. By the interpretation of in silico data combined with in vivo localizations, our results provide indications on protein transport and folding, protein degradation and processing, plastid division, lipid transfer, structural maintenance as well as metabolism in the PPC of *P. tricornutum*. Future studies should deal with a precise characterization of the proteins present in the PPC via biochemical and interaction assays. The combined data set of PPC-located proteins generated in this and previous studies may be of high relevance for the generation of algorithms to identify the soluble proteome of the PPC of *P. tricornutum* in silico. Of further interest are proteomic and targeting studies focusing on membrane proteins that can address specific processes such as energy supply of the PPC or communication pathways between host and symbiont cytoplasms.

Supplementary Material

Supplementary table S1 and supporting information are available at *Genome Biology and Evolution online* (http://gbe.oxfordjournals.org/).

Acknowledgments

This work is dedicated in memory of Prof Dr Werner Wehrmeyer. This work was supported by the LOEWE program of the state of Hessen. D.M. is a fellow of the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology, Marburg. We thank the members of our laboratory for critical discussions.

Literature Cited

Agrawal S, van Dooren GG, Beatty WL, Striepen B. 2009. Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. J Biol Chem. 284:33683–33691.

Aker J, et al. 2007. In vivo hexamerization and characterization of the Arabidopsis AAA ATPase CDC48A complex using forster resonance

energy transfer-fluorescence lifetime imaging microscopy and fluorescence correlation spectroscopy. Plant Physiol. 145:339–350.
Allen MD, Buchberger A, Bycroft M. 2006. The PUB domain functions as a p97 binding module in human peptide N-glycanase. J Biol Chem. 281:25502–25508.
Archibald JM. 2009. The puzzle of plastid evolution. Curr Biol. 19:R81–R88.
Beskow A, et al. 2009. A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation. J Mol Biol. 394:732–746.
Bolte K, et al. 2009. Protein targeting into secondary plastids. J Eukaryot Microbiol. 56:9–15.
Bullmann L, et al. 2010. Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. J Biol Chem. 285:6848–6856.
Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot 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. 358:109–133 discussion 133–134.
Douglas S, et al. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410:1091–1096.
Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW. 2003. ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. Proc Natl Acad Sci U S A. 100:4328–4333.
Gibbs SP. 1979. The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. J Cell Sci. 35:253–266.
Gilson PR, et al. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci U S A. 103:9566–9571.
Gould SB, et al. 2006a. Protein targeting into the complex plastid of cryptophytes. J Mol Evol. 62:674–681.
Gould SB, 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.
Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. Annu Rev Plant Biol. 59:491–517.
Gruber A, et al. 2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol. 64:519–530.
Gruber A, Weber T, Bartulos CR, Vugrinec S, Kroth PG. 2009. Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. J Basic Microbiol. 49:58–72.
Hempel F, et al. 2007. Transport of nuclear-encoded proteins into secondarily evolved plastids. Biol Chem. 388:899–906.
Hempel F, Bullmann L, Lau J, Zauner S, Maier UG. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Mol Biol Evol. 26:1781–1790.
Hempel F, Felsner G, Maier UG. 2010. New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. Mol Microbiol. 76:793–801.
Kalanon M, Tonkin CJ, McFadden GI. 2009. Characterization of two putative protein translocation components in the apicoplast of Plasmodium falciparum. Eukaryot Cell. 8:1146–1154.
Keeling PJ. 2009. Chromalveolates and the evolution of plastids by secondary endosymbiosis. J Eukaryot Microbiol. 56:1–8.
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.
Kitao Y, Harada H, Matsuda Y. 2008. Localization and targeting mechanisms of two chloroplastic beta-carbonic anhydrases in the marine diatom Phaeodactylum tricornutum. Physiol Plant. 133:68–77.
Kroth PG, et al. 2008. A model for carbohydrate metabolism in the diatom Phaeodactylum tricornutum deduced from comparative whole genome analysis. PLoS One. 3:e1426.
Lane CE, et al. 2007. Nucleomorph genome of Hemiselmis andersenii reveals complete intron loss and compaction as a driver of protein structure and function. Proc Natl Acad Sci U S A. 104:19908–19913.
Madsen L, Seeger M, Semple CA, Hartmann-Petersen R. 2009. New ATPase regulators—p97 goes to the PUB. Int J Biochem Cell Biol. 41:2380–2388.
Maier UG, Douglas SE, Cavalier-Smith T. 2000. The nucleomorph genomes of cryptophytes and chlorarachniophytes. Protist 151:103–109.
Mousley CJ, Tyeryar KR, Vincent-Pope P, Bankaitis VA. 2007. The Sec14-superfamily and the regulatory interface between phospholipid metabolism and membrane trafficking. Biochim Biophys Acta. 1771:727–736.
Raven JA. 2010. Inorganic carbon acquisition by eukaryotic algae: four current questions. Photosynth Res. 106:123–134.
Roberts SB, Lane TW, Morel FMM. 1997. Carbonic anhydrase in the marine diatom Thalassiosira weissflogii (Bacillariophyceae). J Phycol. 33:845–850.
Sommer MS, et al. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol Biol Evol. 24:918–928.
Spork S, et al. 2009. An unusual ERAD-like complex is targeted to the apicoplast of Plasmodium falciparum. Eukaryot Cell. 8:1134–1145.
Suzuki T, Park H, Till EA, Lennarz WJ. 2001. The PUB domain: a putative protein-protein interaction domain implicated in the ubiquitin-proteasome pathway. Biochem Biophys Res Commun. 287:1083–1087.
Szabo E, Colman B. 2007. Isolation and characterization of carbonic anhydrases from the marine diatom Phaeodactylum tricornutum. Physiol Plant. 129:484–492.
Tachibana M, et al. 2011. Localization of putative carbonic anhydrases in two marine diatoms, Phaeodactylum tricornutum and Thalassiosira pseudonana. Photosynth Res. doi: 10.1007/s11120-011-9634-4.
Tanaka Y, Nakatsuma D, Harada H, Ishida M, Matsuda Y. 2005. Localization of soluble beta-carbonic anhydrase in the marine diatom Phaeodactylum tricornutum. Sorting to the chloroplast and cluster formation on the girdle lamellae. Plant Physiol. 138:207–217.
van Dooren GG, et al. 2009. A novel dynamin-related protein has been recruited for apicoplast fission in Toxoplasma gondii. Curr Biol. 19:267–276.
Weber T, Gruber A, Kroth PG. 2009. The presence and localization of thioredoxins in diatoms, unicellular algae of secondary endosymbiotic origin. Mol Plant. 2:468–477.
Yang Y, Glynn JM, Olson BJ, Schmitz AJ, Osteryoung KW. 2008. Plastid division: across time and space. Curr Opin Plant Biol. 11:577–584.
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.

Associate editor: Bill Martin