Distribution of the SELMA Translocon in Secondary Plastids of Red Algal Origin and Predicted Uncoupling of Ubiquitin-Dependent Translocation from Degradation

Simone Stork, a Daniel Moog, a Jude M. Przyborski, b Ilka Wilhelmi, a Stefan Zauner, a and Uwe G. Maier a,c

Laboratory for Cell Biology, Philipps-University Marburg, Marburg, Germany; Laboratory for Parasitology, Philipps-University Marburg, Marburg, Germany; and LOEWE-Zentrum für Synthetische Mikrobiologie (SynMikro), Marburg, Germany

Protein import into complex plastids of red algal origin is a multistep process including translocons of different evolutionary origins. The symbiont-derived ERAD-like machinery (SELMA), shown to be of red algal origin, is proposed to be the transport system for preprotein import across the periplastidal membrane of heterokontophytes, haptophytes, cryptophytes, and apicomplexans. In contrast to the canonical endoplasmic reticulum-associated degradation (ERAD) system, SELMA translocation is suggested to be uncoupled from proteasomal degradation. We investigated the distribution of known and newly identified SELMA components in organisms with complex plastids of red algal origin by intensive data mining, thereby defining a set of core components present in all examined organisms. These include putative pore-forming components, a ubiquitylation machinery, as well as a Cdc48 complex. Furthermore, the set of known 20S proteasomal components in the periplastidal compartment (PPC) of diatoms was expanded. These newly identified putative SELMA components, as well as proteasomal subunits, were in vivo localized as PPC proteins in the diatom Phaeodactylum tricornutum. The presented data allow us to speculate about the specific features of SELMA translocation in contrast to the canonical ERAD system, especially the uncoupling of translocation from degradation.

Organelles such as plastids, including those of secondary origin, almost completely rely on protein import from the host cytosol (46, 65). The structure of complex plastids, surrounded by three or four membranes required, in contrast to primary plastids, the evolution of several additional protein transport mechanisms. Complex plastids arose through secondary endosymbiosis, a process which describes the engulfment of a former free-living eukaryotic alga into a eukaryotic host cell (32, 33). During evolution, the symbiont was subsequently reduced in terms of compartmentalization and genome size to an organelle strictly dependent on the host cell (16, 32). Different types of secondary plastids exist in a very broad range of algae and protists, which can be distinguished based on their evolutionary origin (e.g., a red or green alga derived symbiont), as well as on the amount of cellular reduction inside the host cell. Our understanding of the evolution of organisms harboring a secondary plastid of red algal origin has changed in the last few years. According to the chromalveolate hypothesis, six major lineages were grouped together to be of monophyletic origin: cryptophytes, haptophytes, heterokontophytes, peridinin-containing dinoflagellates, apicomplexans, and the non-plastid-containing ciliates, as well as several smaller lineages related to some chromalveolate members (15, 41). However, recent phylogenetic analyses have given rise to extended theories about the evolution of the lineages with a red algal endosymbiont, including serial endosymbiotic events with secondary, as well as tertiary, endosymbioses (21, 22, 26, 27, 56, 61, 71, 75).

It has been shown that the lineages with an endosymbiont of red algal origin share common plastid protein import mechanisms despite remarkable differences resulting from specific features in plastid ultrastructure (10, 37, 65). Import into complex plastids starts cotranslationally at the endoplasmic reticulum (ER) membrane where nascent precursor proteins are synthesized into the ER lumen. This transport step requires a canonical N-terminal signal peptide (SP). In heterokontophytes, cryptophytes, and haptophytes, the outermost plastid membrane, termed the chloroplast ER (cER) membrane, is continuous with the endomembrane system of the host cell; therefore, the Sec61-mediated import already represents transport across the first membrane of complex plastids. In contrast, the plastids of apicomplexans and peridinin-containing dinoflagellates are not connected to the endomembrane system. Thus, after import into the ER lumen, proteins are likely to be transported to the plastid via vesicle transport mechanisms directly from the ER or via the Golgi apparatus (47, 57).

After the preprotein has entered the cER lumen, the SP is thought to be cleaved off, and a transit peptide-like sequence (TPL) is exposed at the new N terminus. The TPL is required for further transport into the periplastidal compartment (PPC), which resembles the naturally reduced cytoplasm of the endosymbiont, and further into the stroma of the plastid. Such transit peptide-like sequences thereby fulfill an additional function in contrast to transit peptides (TP) in primary plastids. Detailed characterization of the TPL revealed a difference between stroma- and PPC-localized proteins. Stromal proteins possess an aromatic (Phe, Tyr, and Trp) or bulky (Leu) amino acid at the +1 position of their TPL, in contrast to PPC proteins (4, 31, 34, 42). However, the observed AXA-FAP motif at the transition between SP and...
the degree of factor conservation and identify main components including recently published full genome sequences, for SELMA heterokontophytes and apicomplexan parasites. In particular, we in organisms with a red algal endosymbiont with focus on five con at the outer membrane of chloroplasts (TOC) (1, 10, 13, 73).

For transport across the second outermost membrane, the periplastidal membrane, a translocon model was proposed to consist of a recycled ER-associated degradation (ERAD) machinery of symbiont origin (67). Support for this model came from the detection of symbiont-specific ERAD components encoded on the nucleomorph of the cryptophyte Guillardia theta, this being the remnant nucleus of the former red algal endosymbiont in the cryptophytes PPC (23). The canonical ERAD removes aberrant or misfolded luminal (ERAD-L) and membrane (ERAD-M and ERAD-C) ER proteins and tags them after retrotranslocation into the cytosol with polyubiquitin moieties for subsequent proteasomal degradation (7, 40, 66).

However, in the symbiont-specific ERAD-like pathway (SELMa) the retrotranslocation machinery of ERAD-L is postulated to be maintained and possesses the capacity to transport proteins from an ER luminal compartment into a cytoplasmic compartment, the PPC. This process is supposed to be uncoupled from degradation. SELMA is conserved in all secondary evolved organisms with a red algal endosymbiont, for which genomic data are available (26, 67, 68). Proteins of the derlin family are still controversially discussed elements of the ERAD-specific translocon. In the diatom Phaeodactylum tricornutum, two symbiont-localized derlins (PtdDer1-1/PtdDer1-2) are expressed which form hetero-oligomers as well as homo-oligomers and show interaction with transit peptide-like sequences of PPC-localized proteins (38). These components are indeed involved in the transport of proteins into the plastid as indicated by a conditional knock-down mutant of the Toxoplasma gondii sDer1 protein which showed impairment in plastid protein import (2).

The translocation process is predicted to be dependent on ubiquitylation, further supported by the presence of a set of ubiquitylation enzymes (39, 67). Additional factors proposed to be involved in SELMA are a symbiont-specific Cdc48 AAA-ATPase with its cofactor Ufd1 and adaptor proteins (55, 67). After translocation, the precursor proteins are likely to undergo deubiquitylation and are either passed on to the translocon in the third outermost membrane or folded in the PPC (13, 39, 55). Although a residual set of 20S proteasomal components was identified in the PPC of diatoms, there is currently no link between SELMA and proteasomal degradation (55).

Having passed through the PPC, transport across the innermost plastid membranes seems to be comparable to primary plastids with a translocon at the inner membrane of chloroplasts (TIC) and a recently identified Omp85 protein which belongs to the family of Toc75 proteins, the core components of the translocon at the outer membrane of chloroplasts (TOC) (1, 10, 13, 73).

Here, we present an update on the SELMA translocation model in organisms with a red algal endosymbiont with focus on five heterokontophytes and apicomplexan parasites. In particular, we mined the genomes of organisms that carry secondary plastids, including recently published full genome sequences, for SELMA proteins. With this collected data set one would expect to define the degree of factor conservation and identify main components of the SELMA system which evolved to function in protein transport at a plastid membrane. Our results are compared to the respective host ERAD system, as well as to red algal ERAD components, from which SELMA originated. In addition, four new PPC-localized proteins similar to factors involved in ERAD could be identified in the diatom P. tricornutum. We also extended the set of core proteasomal components in the PPC of heterokontophytes and discuss their putative function in relation to SELMA.

MATERIALS AND METHODS

Bioinformatic analysis. The protein sequences of ERAD and SELMA, as well as proteasomal components, were collected from published data or retrieved via blastp and tblastn searches. As queries, sequences from the Saccharomyces cerevisiae ERAD system and the P. tricornutum SELMA system were used to search the genomic databases for Phaeodactylum tricornutum v2.0 (12), Thalassiosira pseudonana (5), Fragilariopsis cylindrus (http://genome.jgi-psf.org/Fracy1/Fracy1.home.html), Aureococcus anophagefferens (30), Emilinia huxleyi CCMIP1516 main genome assembly v1.0 (http://genome.jgi-psf.org/Emilh1/Emilh1.home.html), and Guillardia theta CCMIP2712 v1.0 (http://genome.jgi-psf.org/Guith1/ Guith1.home.html). Sequences from Ectocarpus siliculosus (19) and Babesia bovis were searched at the National Center for Biotechnology Information (NCBI) server (http://www.ncbi.nlm.nih.gov/guide/). Apicomplexan sequences were retrieved from the Plasmodium Genomics Resource version 9.0 (6) for Plasmodium, the Toxoplasma Genomics Resource v7.2 (29) for T. gondii and Neospora caninum, TrpavDB version 1.0 (74) and the NCBI server (http://www.ncbi.nlm.nih.gov/guide/) for Theileria parva, and the Cryptosporidium Genomics Resource v4.6 (36) for Cryptosporidium parvum. ERAD sequences for red algae were either retrieved from the genome projects of Cyanidioschyzon merolae (53) and Galdieria sulphuraria (Michigan State University Galdieria Database [http://genomics.msu.edu/galdieria/about.html]) or by local BLAST (blast-2.2.10-ia32-win32) using expressed sequence tags (EST) of Porphyridium cruentum and partial genome data from Callithrix tuberculatum (http://dbdata.rutgers.edu/data/plantae/) generated by Chan et al. (17).

In general, a minimal e-value of 1e-04 was set as threshold for the identification of ERAD/SELMa components on the protein level. However, in cases of weak query sequence significance, matches with a lower e-value were also inspected. In addition, criteria such as domain structure and composition similarity (NCBI Conserved Domain search) were applied for the identification of relevant proteins (51). For proteasomal components, all S. cerevisiae 20S protein sequences were used as queries to collect a data set of putative proteasomal components, which were then classified according to the NCBI Conserved Domain Database (51), which differs from the S. cerevisiae nomenclature (detailed information on different classifications can be found in reference 60).

All gene models of the identified proteins were aligned to genomic and EST sequences, if available. Thereby, missing N and C termini were identified by searching for putative start and stop codons in frame, respectively. If possible, intron borders of the gene models were checked to be in agreement with EST data. The protein sequences were additionally examined for N-terminal targetting sequences to discriminate symbiont proteins from host factors. PPC directed proteins are characterized by the presence of a SP and a TPL. The SignalP 3.0 Server (24) was used for the prediction of a SP with a cutoff of >0.5 by the HMM algorithm. The sequences were then analyzed with the TargetP 1.1 Server (25) with default settings to define the SP as a secretory signal sequence and exclude mitochondrial targeting. In general, the TPL of PPC (symbiont) proteins cannot be predicted accurately with available tools. For this reason, besides performing the prediction with the TargetP 1.1 Server (25) using signal peptide truncated sequences in “plant” mode, the criteria defined in reference 55 were applied. In some cases, a protein model was identified with high similarity to a known symbiont protein of the diatom P. tricornutum or the apicomplexan parasite P. falceparum but without SP prediction. This can be caused by an incorrect gene model prediction due to the lack of EST data or the presence of several putative start codons. There-
fore, these proteins were assigned as symbiont but marked to lack a signal peptide prediction.

Analyses of transmembrane-spanning regions were performed with TOPCONS (9); analyses of domain and coiled-coil prediction were done using SMART (45). Protein sequence alignments were performed with GENEDOC Software (version 2.6.002 [http://www.psc.edu/biomed/genedoc]).

Plasmid construction and transfection of *P. tricornutum*. The predicted PPC proteins were cloned and transfected into the diatom *P. tricornutum*. The sequences of genes containing introns or without EST support were amplified from cDNA, the rest from gDNA, cloned in front of *egfp* into *P. tricornutum* transfection vectors. *ptsu*β, *ptsr*β1, *ptsu*βq were cloned into the nitrate-inducible pPha-NR vector (GenBank accession no. JN180663), *pts*γ1, *pts*β1, *pts*α3, *ptb*β7, and *pthr*π10 into the light-inducible pPha-T1 vector (GenBank accession no. AF219942). For further information about the sequences of in vivo-localized proteins, as well as the primer sequences, see File S1 in the supplemental material. Biolistic transfection into *P. tricornutum* cells was performed as described previously (67, 77). Positive transformants were cultured under standard conditions as described before (3) with 1.5 mM NH₄Cl in permanent cultivation. Protein expression under the control of the nitrate reductase promoter (pPha-NR vector) was induced by cultivation on 0.9 mM NO₃⁻ for 2 days.

Fluorescence microscopy. *P. tricornutum* transformants were fixed with 4% formaldehyde–0.0075% glutaraldehyde in 1× phosphate-buffered saline buffer and analyzed with a confocal laser scanning microscope Leica TCS SP2 using a HCX PL APO 40×/1.25 to 0.75 oil CS objective lens. The fluorescence of enhanced green fluorescent protein (eGFP) and chlorophyll was excited with an argon laser at 488 nm and detected with two photomultiplier tubes at a bandwidths of 500 to 520 nm and 625 to 720 nm for eGFP and chlorophyll fluorescence, respectively.

RESULTS

Identification of ERAD and SELMA components in red algae and organisms with a red algal endosymbiont. In order to identify new ERAD and SELMA components, all available genomic sequences of red algae and organisms with a red algal endosymbiont were screened via BLAST search with queries from the best-studied ERAD system of *Saccharomyces cerevisiae* (40, 66). The recently published genomes of heterokontophytes (the diatom *Fragilariopsis cylindrus*, the brown alga *Ectocarpus siliculosus*, the harmful alga *Aureococcus anophagefferens*), the nuclear genome of the cryptophyte *Guillardia theta*, and the apicomplexan *Neospora* were included in these analyses. Because SELMA was shown to be phylogenetically derived from the ERAD system of the red algal endosymbiont (26), we also included sequences from the red alga * Cyanidioschyzon merolae*, *Porphyridium cruentum*, *Calliathron tuberculosum*, and *Galdieria sulphuraria* in our analyses (see Materials and Methods for a detailed description of the genome data used). In contrast to the other chromalveolate groups, dinoflagellate plastids have only three surrounding membranes, and very little is known about the mechanisms that transport proteins across these membranes (10, 65). Due to the paucity of genomic data for these organisms, we have not included peridinin-containing dinoflagellates in the present study.

We identified genes for conserved ERAD components in all investigated red algal genomes (Table 1). However, due to incomplete data for *Porphyridium cruentum* (EST data) and especially *Calliathron tuberculosum* (partial genome data), only a subset of ERAD factors could be identified. The collected data set for red algae implicates that the progenitor from which the SELMA machinery originated was capable of ERAD-L via the Hrd1 complex, as well as ERAD-C, via the Doa10 complex in the ER membrane. In addition, all proteins required for ubiquitylation and efficient proteasomal substrate delivery after ERAD retrotranslocation are present in red algae.

In organisms with a red algal endosymbiont, the SELMA system exists in parallel with the host ERAD machinery. The discrimination between proteins of both systems is based on the targeting signal of the PPC localized SELMA proteins in contrast to the mostly cytosolic ERAD components (see Materials and Methods). Identification of a SELMA protein is more reliable if a respective host protein with the same putative function can be found. Therefore, a detailed analysis of the host ERAD system of the investigated organisms was included, and almost all ERAD proteins known from *S. cerevisiae* could be identified in the genomes (Table 1; for detailed information, see File S2 and Table S1 in the supplemental material). All organisms encode for the ER membrane proteins Sec61α, Hrd1, the derlin proteins and, with the exception of apicomplexans, also for Doa10. In addition, a cytosolic ubiquitylation machinery, the Cdc48 complex with its cofactors (Npl4 and Ufd1) and all proteasomal substrate delivery factors (Rad23, Dsk2, Png1) were identified.

Our inspection of the SELMA system in secondary evolved algae and apicomplexan parasites showed a high degree of conserved components for this putative protein translocation machinery (Table 1). However, there can be different reasons for failures in the identification of certain proteins. If a protein is present in most of the organisms of one group but lacking in one specific organism, this is likely caused by incomplete genome sequencing and assembly or by incorrect protein model prediction (e.g., for *Aureococcus anophagefferens*). In contrast, a protein not identified in a whole group of organisms may have been lost completely during evolution. Haptophytes and cryptophytes are represented only by one organism, hindering a final conclusion about the presence or absence of specific proteins but allowing considerations of whole protein complexes. The analysis of the newly available genome sequence of the cryptophyte *G. theta* shows that the partially nucleomorph-encoded SELMA system is supplemented with nucleus-encoded factors (Table 1).

Interestingly, from the three ER membrane protein classes—Sec61α, derlin proteins, and the ubiquitin ligase Hrd1, which are discussed as putative ERAD channel proteins—only the derlin proteins are found in the complex plastids of these organisms as membrane proteins with several transmembrane domains, with the exception of a nucleomorph-encoded Hrd1 in cryptophytes (see below). In contrast to derlins, two symbiotic representatives are present in heterokontophytes, haptophytes, and cryptophytes, as is the case for yeast (ScDer1p and ScDfm1p). Ubiquitylation requires a cascade of three enzymes starting with a ubiquitin-activating enzyme (Uba1) which is present in all organisms. At least one symbiont ubiquitin-conjugating enzyme (sUbc) can also be found, but not all putative PPC-targeted sUbc proteins can be assigned to the same *S. cerevisiae* Ubc protein. While heterokontophytes share a sUbc similar to ScUbc6p and at least one other sUbc protein, apicomplexans seem to encode only for one sUbc protein with the highest similarity to ScUbc4p. The ubiquitin ligase sHrd1 of heterokontophytes differs in protein structure from the symbiont ubiquitin ligase of cryptophytes. Several transmembrane domains are predicted for the GisHrd1 protein. Therefore, it more resembles the yeast ScHrd1p structure than the heterokontophyte E3 ligase which contains only one predicted transmembrane domain.
### TABLE 1

Overview of all identified host ERAD and symbiont SELMA components in organisms with secondary plastids of red algal origin compared to ERAD proteins of four red algal species.

| Protein complex/function | Protein name host (symbiont) | Red algae | Haplosporida | Cryptophytes | Apicomplexans |
|-------------------------|-----------------------------|-----------|-------------|--------------|--------------|
| ER translocon Sec61    | X                           | X         | X           | X            | X            |
| Derlin proteins        | X                           | X         | X           | X            | X            |
| Cdc48 complex          | X                           | X         | X           | X            | X            |
| Processing             | X                           | X         | X           | X            | X            |
| Cdc48                 | X                           | X         | X           | X            | X            |
| Ufd1                  | X                           | X         | X           | X            | X            |
| Npl4                  | X                           | X         | X           | X            | X            |
| Doa10                 | X                           | X         | X           | X            | X            |
| Polyubiquitin          | X                           | X         | X           | X            | X            |
| Cdc48-2                | X                           | X         | X           | X            | X            |
| Ufd1-1 (sUfd1)         | X                           | X         | X           | X            | X            |
| Ufd1-2 (sUfd1-2)       | X                           | X         | X           | X            | X            |
| Hsp70                 | X                           | X         | X           | X            | X            |
| Hsp40                 | X                           | X         | X           | X            | X            |

Proteins similar to ScUsa1p, ScCue1p, and ScUbx4p could not be identified in the host or symbiont version and are therefore not included in the table. UK proteins are further assigned to the most similar S. cerevisiae UK enzyme in the supplemental material (see Table S1 in the supplemental material). X, detected; (X), symbiont gene detected by homology but without targeting sequence; a, more than one gene detected. Superscript numbers in parentheses indicate the corresponding reference(s). Protein identifiers can be found in File S2 (Table S1) in the supplemental material.

Dec 2012 Volume 11 Number 12 ec.asm.org
The symbiont Cdc48-complex together with sUfd1 can be found in all organisms investigated, and we were now able to identify a sNpl4 protein in the diatom *P. tricornutum* which is conserved among heterokontophytes. The same is the case for three other newly identified putative symbiont proteins, sUBX, sUbq, and sPng1. These share similarity to ERAD factors and are present in addition to the host version in the diatom *P. tricornutum* and other heterokontophytes. None of these proteins is conserved in apicomplexans (Table 1).

**Newly identified putative SELMA components:** the sCdc48 cofactor sNpl4, a UBX domain-containing protein, a symbiont ubiquilin-like protein and a peptide N-glycanase. Among the newly identified putative SELMA components are two Cdc48 binding proteins, the UBX domain-containing protein sUBX (symbiont UBX) and the Ufd1 cofactor sNpl4, as well as proteins with sequence similarity to the de-glycosylation enzyme ScPng1p (sPng1) and the polyubiquitin-binding protein ScDsk2p (sUbq). Not all four proteins are predicted to have a TPL in the diatom *P. tricornutum*. In such a case, a signal peptide on usually cytosolic proteins is indicative for a PPC localization but remained to be verified in localization experiments. Therefore, PtsNpl4, PtsUBX, PtsUbq, and PtsPng1 were expressed as eGFP fusions in *P. tricornutum* (scale, 10 μm; TL, transmission light; PAF, plastid autofluorescence). (B) Structural overview of the domain composition of PtsNpl4, PtsUbq, and PtsPng1 in comparison to the respective *S. cerevisiae* ERAD protein ScNpl4p, ScDsk2p and ScPng1p. Whereas PtsNpl4 and PtsPng1 share the conserved domains of their yeast counterparts, PtsUbq lacks the UBA domain for polyubiquitin binding. PtsUBX cannot be assigned to a specific ERAD protein. Red, signal peptide; blue, TPL predicted; light blue, no TPL predicted; UBQ, ubiquitin homologues; STI1, heat shock chaperonin-binding motif; UBA, ubiquitin associated domain; TG, transglutaminase/protease-like homologues; Rad4; Rad4 transglutaminase-like domain.

**FIG 1** In vivo localization and domain organization of new SELMA proteins of the diatom *P. tricornutum*. (A) In vivo localizations of PtsUBX, PtsNpl4, PtsPng1, and PtsUbq as eGFP fusion proteins in *P. tricornutum* show the characteristic PPC fluorescence in the middle of the two plastid lobes (scale, 10 μm; TL, transmission light; PAF, plastid autofluorescence). (B) Structural overview of the domain composition of PtsNpl4, PtsUbq, and PtsPng1 in comparison to the respective *S. cerevisiae* ERAD protein ScNpl4p, ScDsk2p and ScPng1p. Whereas PtsNpl4 and PtsPng1 share the conserved domains of their yeast counterparts, PtsUbq lacks the UBA domain for polyubiquitin binding. PtsUBX cannot be assigned to a specific ERAD protein. Red, signal peptide; blue, TPL predicted; light blue, no TPL predicted; UBQ, ubiquitin homologues; STI1, heat shock chaperonin-binding motif; UBA, ubiquitin associated domain; TG, transglutaminase/protease-like homologues; Rad4; Rad4 transglutaminase-like domain.
toms (68). Unfortunately, it was not possible to detect a symbiont ubiquitin in the other newly investigated organisms. Interestingly, the position Lys63, usually used for polyubiquitin linkages related to modifications of protein function, is also no longer present in all symbiont ubiquitins except the second domain of GtsUbi.

Identification and localization of proteasomal components in the PPC of heterokontophytes and cryptophytes. We previously reported on the presence of relict 20S proteasomal components in the \textit{P. tricornutum} PPC (55). Here, we expand the model of a symbiont core proteasome by extensive \textit{in silico} analyses.

The cryptophyte \textit{Guillardia theta}, which is still able to synthesize proteins in the PPC, encodes for an almost complete set of PPC-localized proteasomal degradation components on its nucleomorph genome (23). This is in contrast to other organisms with a red algal endosymbiont and, among them, only in heterokontophytes could residual 20S subunits be identified. Importantly, we did not detect a symbiont 19S regulatory particle in the PPC of heterokontophytes, although it is present in cryptophytes. We could not identify a complete set of 20S subunits, including seven alpha and seven beta subunits, for any of the organisms studied (Table 2). Instead, the putative 20S core particle seems to vary in subunit composition, with the exception of conserved $s_{3}/H9251$, $s_{3}/H9252$, and $s_{3}/H9251$.

TABLE 2

| Protein | C. merolae | C. reinhardtii | C. sulphuraria | \textit{P. tricornutum} | \textit{T. pseudonana} | \textit{F. cylindrus} | \textit{E. siliculosus} | \textit{A. amphibiguum} | \textit{G. theta} |
|---------|------------|---------------|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|
| $\alpha 1$ | X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\alpha 2$ | X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\alpha 3$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\alpha 4$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\alpha 5$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\alpha 6$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\alpha 7$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| (general) | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 1$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 2$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 3$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 4$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 5$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 6$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |
| $\beta 7$ | X X X X X $X^{(55)}$ | X X X X | $X^{(55)}$ | X X X X | X X X X | X X X X | X X X X | X X X X | X X X X |

$^a$X, detected; $^b$more than one gene detected; (X), symbiont gene detected by homology but without targeting sequence. Superscript numbers in parentheses indicate the corresponding reference. Protein identifiers can be found in File S2 (Table S2) in the supplemental material.
fusion with a 5′-3′ exonuclease, also lacking a signal peptide. In A. anophagefferens and E. siliculosus in addition to a symbiont s5 with signal peptide prediction, several putative symbiont subunits exist but an exact defining of the gene model is difficult. Therefore, a classification into host or symbiont protein cannot yet conclusively be determined.

In addition to the already-reported symbiont 20S proteasomal subunits Ptsβ2, Ptsβ6, Ptsβ7, Ptsα7-1, and Ptsα7-2 (55), we successfully localized two additional subunits in the PPC of P. tricornutum. Both Ptsα3-1 and Ptsβ1 showed the typical PPC fluorescence pattern (Fig. 3). As a comparison, two subunits of the host proteasome were also localized, Pthβ7 and PthRpn10, which resulted in a different fluorescence pattern outside of the plastid.

DISCUSSION

For protein transport across the periplastidal membrane of complex plastids of red algal origin, an ERAD-derived mechanism (SELMA) was proposed as the protein translocation machinery (67). The SELMA model, originally based on our findings in cryptophytes, was shown to be conserved in organisms with a red algal endosymbiont (2, 26, 67, 68), and SELMA components can be identified in all available genomes in addition to the host ERAD machinery (Table 1). However, the exact mechanism of this transport step, the pore-forming proteins, and the minimal required components remain an open question. The proposed SELMA components show often a minimized structure, as domains, known from ERAD proteins of other organisms, are missing. Thus, the SELMA complex should indicate a minimized version of the retro-translocation activity of ERAD in general. Different extents of ERAD to SELMA reduction can be found in the investigated organisms according to the amount of reduction of the former endosymbiont. The cryptophyte G. theta represents the most extended set of SELMA and proteasomal components, resulting most probably from its transcriptionally and translationally active nucleomorph in the PPC. Apicomplexans instead have the smallest set of identified SELMA components and seem to lack a symbiont proteasome. Thus far, all investigated organisms share the following as SELMA components in the PPC: derlins as membrane and putative channel proteins, a ubiquitylation machinery, and a Cdc48 complex. It remains to be determined whether the recently identified conserved PPC protein PPP1 provides a new crucial function for protein transport (64). All other identified proteins with functions related to SELMA or the proteasome in heterokontophytes might represent lineage specific adaptations (Fig. 4).

Of all ER membrane proteins that are candidates for a translocation channel in ERAD (66), only the derlin proteins could be identified as SELMA components. Apart from that, the Sec61...
channel and the ubiquitin ligase Hrd1 are in discussion as potentially being capable of fulfilling this function. On the one hand, we could not identify additional symbiont Sec61 subunits; on the other hand, the diatom symbiont E3 ubiquitin ligase is predicted to have only one transmembrane domain and is therefore most likely not capable of homotypic channel formation. However, the derlin proteins and the ubiquitin ligase might form a membrane complex which connects translocation to ubiquitylation. The presence of the ubiquitylation enzymes sUba and sUbc and ubiquitin itself in almost all investigated organisms, including apicomplexans, suggests the presence of a ubiquitin-dependent mechanism in the PPC. Once the preprotein is ubiquitylated in the PPC, it can be recognized by the Cdc48 complex (76). We could identify at least one symbiont-specific sCdc48 protein in all organisms investigated. The Cdc48-ATPase has been shown to be a central component of ERAD, acting specifically in concert with its cofactors Ufd1 and Npl4 (54, 76). Although Cdc48 is known to have various cellular functions, the identification and localization of a symbiont Npl4 protein of *P. tricornutum* presented here now defines the sCdc48-sUfd1-sNpl4 complex as a SELMA component. However, other functions unrelated to protein transport together with thus-far-unidentified cofactors cannot be excluded. The new PPC-localized protein PtsPng1, most likely a sCdc48 binding protein due to its UBX domain, might also be involved in SELMA translocation, akin to the case for UBX proteins in ERAD. These proteins can direct the Cdc48-ATPase to a specific protein complex in the context of ERAD to ubiquitin ligases at the ER membrane (50, 62, 63).

After translocation is completed, ERAD substrates are recognized by a set of cytosolic proteins and processed for degradation by the proteasome (59, 76). The presence of a relict symbiont proteasome in heterokontophytes (Table 2) raises the question of a functional link between the ERAD derived SELMA machinery and degradation in the PPC of these organisms. Several features of both machineries in the PPC argue against such a connection. On the one hand, the PPC of apicomplexan parasites and haptophytes lacks symbiont 20S subunits and therefore harbors a SELMA system which seems to be completely independent of a proteasomal function. On the other hand, proteasomal substrates are not only delivered by the ERAD system but can also be degraded independently of ubiquitylation (8). The 20S core particle was shown to have basal proteolytic activity toward unstructured or oxidized proteins (60). It is also implicated in maturation and specific cleavage of various proteins, which gain access to the proteolytic chamber through an interaction with N termini of the α-subunits (8, 49). In the PPC of heterokontophytes, we identified proteasomal subunits of the 20S core particle, including proteolytic active subunits. It was not possible to detect a complete set of 20S subunits in any of the heterokontophyte species. Either the remaining α- and β-subunits in the genomes are too divergent to be recognized or the putative reduced 20S particle in the PPC can vary in subunit composition, replacing some subunits by other ones.

In addition, the recognition and unfolding of ubiquitylated proteasomal substrates is mediated by the 19S regulatory particle of the proteasome, which was not identified in a PPC-targeted version in heterokontophytes. However, the two newly identified PPC proteins sUbq and sPng1 are counterparts to ScDsk2p and ScPng1p, which are known from ERAD to function between retrotranslocation and degradation (43). The symbiont ubiquitin-like protein PtsUbg lacks the C-terminal ubiquitin-associated (UBA) domain for polyubiquitin binding present in ScDsk2p and mammalian ubiquilins for recognition of proteasomal substrates (28, 48). In addition, the ubiquitin-like domain (UBQ) at the N terminus of ScDsk2p was shown to bind proteasomal components of the 19S regulatory particle, as well as to ScUfd2p (35, 69), both not present in a symbiont version in the PPC. Most likely, both proteins had to adapt to new functions. PtsPng1, the PPC-localized peptide N-glycanase, might either be involved in the maturation of PPC-localized proteins or be required for efficient removal of glycan moieties of plastid precursor proteins added in the ER lumen before transport across the third outermost plastid membrane. In contrast to heterokontophytes, haptophytes, and cryptophytes, apicomplexans encode neither a host nor a symbiont Png1 protein. This is likely due to a reduction of N-glycosylation capacities in these organisms, especially for apicoplast proteins (14).

Another important feature of SELMA is the symbiont ubiquitin, which shows alterations at specific lysine residues. In heterokontophytes, the PPC-localized ubiquitin (sUbi) does not possess the conserved lysine residues Lys48 and Lys63, whereas haptophyte and apicomplexan (except *P. falciparum*) sUb sequences still contain Lys48 but show mutations at Lys63. Lys48 was shown to represent the most prominent position for polyubiquitylation leading to proteasomal degradation (72). Although recent work suggests a more complex interplay between different ubiquitin linkages on various lysine residues also in degradation (44), the loss of Lys48 in the symbiont ubiquitins of heterokontophytes might be an evolutionary adaptation required for the establishment of the symbiont ERAD as a preprotein translocation system. This is supported by the finding that only organisms with a symbiont proteasome (Table 2) show this ubiquitin Lys48 modification. An exception is the cryptophyte *G. theta* with two ubiquitin domains in the predicted GtsUbi sequence, one overall conserved and another having Lys mutations at both positions. One might speculate about a separation of SELMA and proteasomal degradation in the cryptophytes PPC based on different ubiquitins, which might be caused by the different morphology, since cryptophytes—in contrast to all other organisms with a secondary red algal symbiont—still synthesize proteins in the PPC. Ubiquitin Lys63 is implicated in ubiquitylation processes related to functional modifications of target proteins (72). Remarkably, loss of Lys63 in all organisms with a secondary plastid of red algal origin leads to reduced ubiquitylation possibilities in the PPC in contrast to the manifold mechanisms regulated by ubiquitylation in the host cytosol. Thus, it remains to be determined whether the symbiont ubiquitins can be used for both mono- and polyubiquitylation on the remaining lysine residues.

The SELMA translocation machinery provides an interesting view into evolutionary rearrangements and modifications of already existing mechanisms. During the establishment of a red alga as an organelle, the symbiont ER-associated degradation machinery was split into a translocation complex on the one hand and a presumed degradation machinery on the other one. The former now represents the second step of protein import into complex plastids across the periplastidal membrane, whereas the latter one might be required for protein homeostasis in the PPC of only certain groups of organisms with a red algal endosymbiont. Such modularization of the well conserved ERAD translocation not only gave rise to SELMA but also showed to be the principle.
mechanism of the peroxisomal importomer, again a ubiquitin-dependent translocation independent of proteasomal degrada-
tion (11).

ACKNOWLEDGMENTS
This study was supported by the Deutsche Forschungsgemeinschaft (Collaborative Research Centre 593 for S.S., J.W., S.Z., and U.-G.M. and SFB TR1 for J.M.P.). D.M. is a fellow of the International Max Planck Research School for Environmental, Cellular and Molecular Microbiol-
ey (IMPRS-MIC).

REFERENCES
1. Agrawal S, Striepen B. 2010. More membranes, more proteins: complex protein import mechanisms into secondary plastos. Protist 161:672–687.
2. Agrawal S, van Dooren GG, Beatty WL, Striepen B. 2009. Genetic evidence that an endosymbiotic-derived endoplasmic reticulum–
associated protein degradation (ERAD) system functions in import of apicoplast proteins. J. Biol. Chem. 284:33683–33691.
3. Apt KE, Grossman A, Kroth-Pancic P. 1996. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. Mol. Gen. Genet. 252:572–
579.
4. Apt KE et al. 2002. In vivo characterization of diatom multipartite plastid targeting signal. J. Cell Sci. 115:4061–4069.
5. Armbrust EV, Berges JA, Bowler C, Green BR. 2004. The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science 306:79–86.
6. Bagola K, Mehnert M, Jarosch E, Sommer T. 2009. Protein targeting into complex diatom plastids: functional characterization of cellular proteins imported by ubiquitin-
transferase. J. Mol. Biol. 386:814–827.
7. Bernsel A, Viklund H, Hennerdal A, Elofsson A. 2009. TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res. 37:W465–468.
8. Bolte K, et al. 2009. Protein targeting into secondary plastids. J. Eukaryot. Microbiol. 56:9–15.
9. Bolte K, et al. 2011. Making new out of old: recycling and modification of an ancient protein translocation system during eukaryotic evolution: mechanistic comparison and phylogenetic analysis of ERAD, SELMA, and the peroxisomal importomer. Bioessays 33:368–376.
10. Bowler C, et al. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.
11. Bullmann L, et al. 2010. Filling the gap: evolutionarily conserved Omp85 in plastids of chromalveolates. J. Biol. Chem. 285:6848–6856.
12. Bushkin GG, et al. 2010. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. 37:D539–D543.
13. Bagola K, Mehrert M, Jarosch E, Sommer T. 2011. Protein dislocation dependent translocation independent of proteasomal degrada-
tion. Proc. Natl. Acad. Sci. U. S. A. 99:745–750.
14. Gajria B, et al. 2008. ToxopDB: an integrated Toxoplasma gondii database resource. Nucleic Acids Res. 36:D553–D556.
15. Uphoff B, et al. 2011. Niche of harmful alga Auroeococcus anophagefferens revealed through ecogenomics. Proc. Natl. Acad. Sci. U. S. A. 108:4352–
4357.
16. Green BR. 2011. After the primary endosymbiosis: an update on the chromalveolates hypothesis and the origins of algae with Chl c. Photosynth. Res. 107:103–115.
17. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier Kroth-U. 2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol. Biol. 64:519–530.
18. Hänzelmann P, Stinger J, Hofmann K, Schindelin H, Rassl S. 2010. The yeast E4 ubiquitin ligase Ufd2 interacts with the ubiquitin-like domains of Rad23 and Dsk2 via a novel and distinct ubiquitin-like binding domain. J. Biol. Chem. 285:20390–20398.
19. Heiges M, et al. 2005. CryptodoB: a Cryptosporidium bioinformatics resource update. Nucleic Acids Res. 33:D419–D422.
20. Hempel F, et al. 2007. Transport of nuclear-encoded proteins into secondary plastids. J. Biol. Chem. 282:889–906.
21. Hempel F, Bullmann L, Lau J, Zauner S, Maier U. 2009. ERAD-derived preprotein transport across the second outermost plastid mem-
brane of diatoms. Mol. Biol. Evol. 26:1781–1790.
22. Hempel F, Felsner G, Maier U. 2010. New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. Mol. Microbiol. 76:793–801.
23. Hoseki J, Ushioda R, Nagata K. 2010. Mechanism and components of endoplasmic reticulum-associated degradation. J. Biochem. 147:19–25.
24. Keeling PJ. 2009. Plastid evolution. Annu. Rev. Plant Biol. 60:745–750.
25. Kravtsova-Ivantsiv Y, Ciechanover A. 2009. New proteins in the apicoplast membrane and the independent evo-
lutionary history of diatom genomes. Trends Plant Sci. 14:617–621.
26. Leupin I, Geier H-G, Fuchs B, Planck S. 2000. Predicting protein domain annotation resource. Nucleic Acids Res. 28:353–366.
27. Lim PJ, et al. 2009. New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. Mol. Microbiol. 76:793–801.
28. Kim I, et al. 2006. The Pml-Rad23 complex regulates glycoprotein turnover. J. Cell Biol. 172:211–219.
29. Krotzova-Ivantsiv Y, Georgacheva A. 2012. Non-canonical ubiquitin- 
based signals for proteasomal degradation. J. Cell Sci. 125:539–548.
30. Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40:D302–D305.
31. Li H-M, Chiu C-C. 2010. Protein transport into chloroplasts. Annu. Rev. Plant Biol. 61:157–180.
32. Lim I, Kalanon M, McFadden GI. 2009. New proteins in the apicoplast membranes: time to rethink apicoplast protein targeting. Trends Parasit-
ology. 25:197–200.
33. Lim IJ, Felsner G, Maier U. 2010. New mechanistic insights into pre-protein transport. J. Cell Sci. 123:2725–2733.
34. Lim IJ, Felsner G, Maier U. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Mol. Biol. Evol. 26:1781–1790.
35. Hathaway J, Stegeman J, Bowler C, Green BR. 2004. Improved protein targeting into complex diatom plastids. Plant J. 41:175–183.
36. Kim I, et al. 2006. The Pmg-I-Rad23 complex regulates glycoconjugate turnover. J. Cell Biol. 172:211–219.
37. Krotzova-Ivantsiv Y, Georgacheva A. 2012. Non-canonical ubiquitin-
based signals for proteasomal degradation. J. Cell Sci. 125:539–548.
38. Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40:D302–D305.
39. Li H-M, Chiu C-C. 2010. Protein transport into chloroplasts. Annu. Rev. Plant Biol. 61:157–180.
40. Lim I, Kalanon M, McFadden GI. 2009. New proteins in the apicoplast membranes: time to rethink apicoplast protein targeting. Trends Parasit-
ology. 25:197–200.
41. Lim IJ, Felsner G, Maier U. 2010. New mechanistic insights into pre-protein transport. J. Cell Sci. 123:2725–2733.
51. Marchler-Bauer A, et al. 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 39:D225–D229.
52. Mason JM, Arndt KM. 2004. Coiled coil domains: stability, specificity, and biological implications. Biochemiochem 3:170–176.
53. Matsuzaki M, et al. 2013. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. Nat. Cell Biol. 14:117–123.
54. Meyer H, Bug M, Bremer S. 2012. The Cdc48 machine in endoplasmic reticulum (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions. FASEB J. 16:635–641.
55. Moog D, Stork S, Zauner S, Maier U-G. 2011. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biol. Evol. 3:375–382. doi:10.1093/gbe/evr031.
56. Moustafa A, et al. 2006. Assembly and function of the proteasome. Cell. Mol. Life Sci. 63:2867–2874.
57. Nassoury N, Cappadocia M, Morse D. 2003. Plastid ultrastructure defines the protein import pathway in dinoflagellates. J. Cell Sci. 116:2867–2874.
58. Patron NJ, Waller RF. 2007. Transit peptide diversity and divergence: a global analysis of plastid targeting signals. Bioessays 29:1048–1058.
59. Richly H, et al. 2005. A series of ubiquitin binding factors connects Cdc48/p97 to substrate multiubiquitylation and proteasomal targeting. Cell 120:73–84.
60. Sasaki Y, Tanaka K. 2012. Assembly and function of the proteasome. Methods Mol. Biol. 832:315–337.
61. Sanchez-Puerta MV, Delwiche CF. 2008. A hypothesis for plastid evolution in chromalveolates. J. Phycol. 44:1097–1107.
62. Schuberth C, Buchberger A. 2005. Membrane-bound Ubx2 recruits Cdc48 to ubiquitin ligases and their substrates to ensure efficient ER-associated protein degradation. Nat. Cell Biol. 7:999–1006.
63. Schuberth C, Buchberger A. 2008. UBX domain proteins: major regulators of the AAA ATPase Cdc48/p97. Cell. Mol. Life Sci. 65:2360–2371.
64. Sheiner L, et al. 2013. A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. PLoS Pathol. 7:e1002392. doi:10.1371/journal.ppat.1002392.
65. Sheiner L, Striepen B. 2012. Protein sorting in complex plastids. Biochim. Biophys. Acta doi:10.1016/j.bbamcr.2012.05.030.
66. Smith MH, Ploegh HL, Weissman JS. 2011. Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. Science 334:1086–1090.
67. Sommer MS, et al. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol. Biol. Evol. 24:918–928.
68. Spork S, et al. 2009. An unusual ERAD-like complex is targeted to the apicoplast of Plasmodium falciparum. Eukaryot. Cell 8:1134–1145.
69. Su V, Lau AF. 2009. Ubiquitin-like and ubiquitin-associated domain proteins: significance in proteasomal degradation. Cell. Mol. Life Sci. 66:2819–2833.
70. Suzuki T, Park H, Lennarz WJ. 2002. Cytoplasmic peptide:N-glycanase (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions. FASEB J. 16:635–641.
71. Teich R, Zauner S, Baurain D, Brinkmann H, Petersen J. 2007. Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. Protist 158:263–276.
72. Trempe J-F. 2011. Reading the ubiquitin postal code. Curr. Opin. Struct. Biol. 21:792–801.
73. van Dooren GG, Tomova C, Agrawal S, Humbel BM, Striepen B. 2008. Toxoplasma gondii Tic20 is essential for apicoplast protein import. Proc. Natl. Acad. Sci. U. S. A. 105:13574–13579.
74. Visendi P, et al. 2011. TparvaDB: a database to support Thelnergia parva vaccine development. Database (Oxford) 2011:bar015.
75. Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related Chromera velia. Genome Biol. Evol. 3:1220–1230. doi:10.1093/gbe/evr100.
76. Wolf DH, Stolz A. 2012. The Cdc48 machine in endoplasmic reticulum associated protein degradation. Biochim. Biophys. Acta 1823:117–124.
77. Zaslavskaya IA, 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.