DOI: 10.1534/genetics.107.085704

The Chloroplast Protein Translocation Complexes of *Chlamydomonas reinhardtii*: A Bioinformatic Comparison of Toc and Tic Components in Plants, Green Algae and Red Algae

Ming Kalanon and Geoffrey I. McFadden[1]

Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, 3010 Victoria, Australia

Manuscript received December 12, 2007
Accepted for publication March 19, 2008

ABSTRACT

The recently completed genome of *Chlamydomonas reinhardtii* was surveyed for components of the chloroplast protein translocation complexes. Putative components were identified using reciprocal BlastP searches with the protein sequences of *Arabidopsis thaliana* as queries. As a comparison, we also surveyed the new genomes of the bryophyte *Physcomitrella patens*, two prasinophyte green algae (*Ostreococcus lucimarinus* and *Ostreococcus tauri*), the red alga *Cyanidioschizon merolae*, and several cyanobacteria. Overall, we found that the components of the import pathway are remarkably well conserved, particularly among the Viridiplantae lineages. Specifically, *C. reinhardtii* contained almost all the components found in *A. thaliana*, with two exceptions. Missing from *C. reinhardtii* are the C-terminal ferredoxin-NADPH-reductase (FNR) binding domain of Tic62 and a full-length, TPR-bearing Toc64. Further, the N-terminal domain of *C. reinhardtii* Toc34 is highly acidic, whereas the analogous region in *C. reinhardtii* Toc159 is not. This reversal of the vascular plant model may explain the similarity of *C. reinhardtii* chloroplast transit peptides to mitochondrial-targeting peptides. Other findings from our genome survey include the absence of Tic22 in both Ostreococcus genomes; the presence of only one Toc75 homolog in *C. merolae*; and, finally, a distinctive propensity for gene duplication in *P. patens*.

THE completion of the *Chlamydomonas reinhardtii* genome presents an opportunity for a genomewide survey of components making up its chloroplast protein translocation complexes. This first glimpse of the Chlamydomonas chloroplast protein import machinery provides an important new perspective on our models of chloroplast protein translocators, which until recently have relied heavily upon studies of vascular plants, particularly *Arabidopsis thaliana* and *Pisum sativum*.

Despite significant variation in plastid morphology and function, all plastids derive from a single endosymbiosis (Martin and Herrmann 1998; Cavalier-Smith 2000; López-Juez 2007), which occurred >930 million years ago (Berney and Pawlowski 2006) and possibly >1200 million years ago (Butterfield 2000). Although the endosymbiont retained its prokaryotic double-membrane architecture along with its thylakoid membranes, it ceded control of the majority of its genetic blueprint, with most of its genome being lost or transferred to the host nucleus (Martin *et al.* 1998). Currently, plastid genomes contain only 50–200 protein-encoding genes, a fraction of the original number of genes that would have been possessed by the cyanobacterium-like endosymbiont (Martin *et al.* 2002; Leister 2003; Timmis *et al.* 2004). However, this loss of genes from nascent plastids was not accompanied by a cognate reduction in plastid metabolic function or activity. Indeed, while plastids have abandoned some activities common to the cyanobacterial forebears, they still practice a diverse retinue of metabolism and contain an estimated 1000–2000 proteins. Most of these plastid proteins are encoded by nuclear genes and imported post-translationally from the cytosol. Thus, one fundamental requirement of plastid evolution is a protein translocation system to facilitate the post-translational return of endosymbiont proteins back to the organelle. It is likely that at least a rudimentary form of such a translocation system existed soon after the initial endosymbiotic event, since the majority of the gene transfer from the endosymbiont had already occurred at this time (Martin *et al.* 1998; Timmis *et al.* 2004).

Our current understanding of plastid protein import complexes stems from two decades of elegant cell biological and genetic studies focused on vascular plants (reviewed in Soll and Schleiff 2004). A host of proteins have been identified as components of two quasi-independent translocons resident in the outer and inner membranes of the plant plastid, known as Toc (*t*ranslocon at the *o*uter *c*hloroplast envelope) and Tic (*t*ranslocon at the *i*nner *c*hloroplast envelope). These translocons act in tandem to transport proteins across the two membranes while maintaining the redox integrity of the organelle.

After plastid acquisition, photosynthetic eukaryotes diverged into three lineages: namely the glaucocysto-

[1]*Corresponding author:* Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, 3010 Victoria, Australia. E-mail: gim@unimelb.edu.au

phytes; the rhodophytes; and the Viridiplantae, comprising the green algae and the land plants. Within the green algae are several major lineages, including the early diverging prasinophytes, the chlorophytes (to which Chlamydomonas belongs), and the charophytes (the sister group to land plants). The Chlamydomonas genome (Merchant *et al.* 2007) provides an opportunity to predict which components a chlorophyte probably uses for chloroplast protein import in comparison with models established in land plants. At the same time, we can extend this perspective to embrace the recently completed genomes of two species of Ostreococcus from the prasinophytes and, wider still, to a member of the red algae (*Cyanidioschizon merolae*), providing the first overview of this key component of plastid evolution in plants, green algae, and red algae.

Using reciprocal BLAST searches with defined plant Toc and Tic orthologs, our aim was to create an *in silico* model of the *C. reinhardtii* chloroplast protein import complexes. In this analysis, we focused mainly on primary plastids, since import into plastids derived from secondary endosymbiosis requires translocation across more than two membranes and involves novel adaptations of different import complexes (Sommer *et al.* 2007). Previous work has also shown that secondary plastids lack a large number of known Toc and Tic homologs (McFadden and Van Dooren 2004).

MATERIALS AND METHODS

With the exception of Tic21, *P. sativum* (NCBI) Toc and Tic components were used to identify orthologs in *A. thaliana* [The Arabidopsis Information Resource (TAIR) protein database, version TAIR7_pep_20070425, http://www.arabidopsis.org], which in turn were used as queries in BlastP searches of the *C. reinhardtii* genome [version 3.1 at the Joint Genome Institute (JGI) http://genome.jgi-psf.org/]. Tic21 was originally identified in *A. thaliana* and has no known *P. sativum* ortholog. Significant results were then used in reciprocal BlastP searches against the *A. thaliana* genome, and the resulting *e*-value, a score indicating the accuracy of the BlastP result, was recorded in Table 1. In a similar way, Toc and Tic components were identified in the genomes of *Physcomitrella patens* (version 1.1 at JGI), *Ostreococcus lucimarinus* (version 2.0 at JGI), *Ostreococcus tauri* (version 2.0 at JGI), and *C. merolae* (version 5.0 at http://merolae.biol.s.u-tokyo.ac.jp), as well as in 10 representative cyanobacterial genomes (*Anabaena variabilis*, ATCC 29413; *Crocosphaera watsonii*, WH 8501; *Gloeobacter violaceus*, PCC 7421; *Nostoc punctiforme*, PCC 73102; *Nostoc* sp., PCC 7120; *Prochlorococcus marinus* str. MIT, 9312; *Synechococcus elongatus*, PCC 6301; *Synechocystis* sp., PCC 6803; *Thermosynechococcus elongatus*, BP-1; and *Trichodesmium erythraeum*, IMS101 at NCBI). Sequences from JGI genomes were accessed from the Genome Browser map to ensure that the most appropriate gene models were analyzed. For the cyanobacterial genomes, only the best-hit result was used in the reciprocal BlastP.

All BlastP searches against the *A. thaliana* genome used TAIR BLAST version 2.2.8. Searches against *C. reinhardtii*, *P. patens*, and *Ostreococcus* spp. used BlastP programs in the respective JGI databases. Searches in *C. merolae* used the BlastP algorithm in the genome site (http://merolae.biol.s.u-tokyo.ac.jp/blast/blast.html). BlastP searches against cyanobacterial genomes were simultaneously performed on the NCBI server (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). BlastP and TBlastN searches of chloroplast genomes utilized the collated chloroplast genome database Chloroplast DB (http://chloroplast.cbio.psu.edu). Sequence logo alignments were constructed with Weblogo (version 2.8.2, http://weblogo.berkeley.edu/logo.cgi; Crooks *et al.* 2004). Subcellular localization predictions were made with the neural network TargetP (version 1.1, http://www.cbs.dtu.dk/services/TargetP; Emanuelsson *et al.* 2007) using plant algorithms. Transmembrane α-helices were predicted with TMHMM (version 2.0, http://www.cbs.dtu.dk/services/TMHMM; Krogh *et al.* 2001), and the combined topology predictor Phobius was also used for Toc12 analysis (http://phobius.sbc.su.se; Käll *et al.* 2004). Transmembrane β-barrel predictions for Tic110 were made with PROFtmb (http://cubic.bioc.columbia.edu/services/proftmb; Bigelow *et al.* 2004). Phylogenetic analysis of Tic20 required MacClade (version 4.06) to define an inclusion set of 125 characters, along with PAUP (version 4.0) to construct preliminary parsimony trees. A maximum-likelihood tree with bootstrap values was constructed with Phyml (Guindon and Gascuel 2003). Protein domains and motifs were identified using two databases: PFAM (version 22.0, http://pfam.sanger.ac.uk) and InterProScan (version 16.1, http://www.ebi.ac.uk/tools/interproscan), using all the available applications.

RESULTS AND DISCUSSION

The *C. reinhardtii* genome was surveyed for chloroplast protein translocation components using reciprocal BlastP homology searches (Table 1). Whereas similar bioinformatic searches utilized only best-hit results of reciprocal Blast (for example, Baum *et al.* 2006; Merchant *et al.* 2007), this conservative approach would have failed to detect several putative orthologs. Several translocation components are represented by closely related paralogs or consist of highly conserved protein domains and motifs, meaning that the *A. thaliana* sequence used in the initial BlastP was not necessarily the best hit in the reciprocal BlastP. By using a combination of reciprocal BlastP and manual curation, our approach reaches a compromise between accuracy and sensitivity of detection.

Toc75, the outer membrane translocation channel: Toc75 is the central translocation pore of the Toc complex (reviewed in Soll and Schleiff 2004). It belongs to the larger prokaryotic Omp85 family of transmembrane β-barrel proteins that include outer membrane porin proteins of gram-negative bacteria (Gentle *et al.* 2005) and the mitochondrial outer membrane proteins Tom40 and Sam50/Tom55 (Kozjak *et al.* 2003; Paschen *et al.* 2003). The *A. thaliana* genome contains four paralogs: AtToc75-III, AtToc75-IV, AtToc75-I, and AtToc75-V/AtOEP80 (Table 1). Of these, AtToc75-III is the functional ortholog of PsToc75 (Baldwin *et al.* 2005).

We detected two putative *C. reinhardtii* Toc75 homologs, of which protein 195512 (CrToc75) is the most orthologous to AtToc75-III (Table 1). CrToc75 contains protein domains typical of Toc75 proteins, including the

TABLE 1

Distribution of Toc and Tic components

	Vascular plants		Moss	Green algae			Red algae	Cyanobacteria
Component	*P. sativum*	*A. thaliana*	*P. patens*	*C. reinhardtii*	*O. lucimarinus*	*O. tauri*	*C. merolae*	
Toc75	Q43715	**At3g46740 (Toc75-III);** At4g09080 (Toc75-IV); *At5g19620 (Toc75-V/OEP80);* *At1g35860 (Toc75-1)*	**171916 (0.0);** 111465 (0.0)*; 160550 (0.0); *180714 (e-109);* *110734 (6e-62)*	**195512 (3e-82);** *195498 (9e-30)*	**24053 (1e-141);** *17254(1e-34)*	**9745 (e-140);** *12791 (1e-34)*	CMJ202C (0.068)	YP_320615 (2e-42) *A. variabilis*
Toc34	Q41009	**At5g05000 (Toc33); At1g02280 (Toc34)**	**211678 (3e-94);** 125298 (2e-89); 113310 (4e-62)	**187290 (9e-36)**	**30720 (4e-52)**	**31832 (1e-51)**	**CMP284C (5e-20)**	*YP_722073 (2e-04) T.erythraeum (GTPase only)*
Toc159	AAF75761	At4g02510 (Toc159); **At2g16640 (Toc132);** At3g16620 (Toc120); At5g20300 (Toc90)	216050 (0.0); 189669 (0.0); **216964 (0.0);** 188734 (0.0)	**206294 (4e-37)**	**25568 (3e-96)**	**37000 (2e-85)**	**CMQ137C (6e-11)**	
Tic110	CAA92823	**At1g06950 (Tic110)**	**228211 (0.0); 177754 (0.0)**	**206003 (4e-38)**	**24868 (6e-79)**	**28497 (1e-133)**	**CMQ342C (9e-25)**	
Tic20	AAC64607	**At1g04940 (Tic20-I);** At4g03320 (Tic20-IV); At2g47840 (Tic20-II); At5g55710 (Tic20-V)	**62632 (3e-80);** 107806 (2e-76); 65298 (1e-04)	**33679 (1e-14);** 183668 (1e-05)*	**33348 (7e-11);** 16587 (0.061); 33082 (5.00)	**34583 (1e-08);** 18958 (1e-04)	**CMS050C (3e-06);** CMV078C (0.010)	NP_488844 (0.001) *N. PCC7120*
Tic21/PIC1	N/A	**At5g15290 (Tic21/CIA5)**	**163267 (2e-63); 170725 (4e-67); 209975(4e-65)**	**113397 (1e-16)*; 178114 (5e-17)**	**16793 (3e-31)**	**19162 (2e-36)**	**CMN128C (2e-22)**	ZP_00105923 (2e-09) *N. punctiforme*
Tic22	AAC64606	**At4g33350 (Tic22-IV);** At3g23710 (Tic22-III)	**171249 (3e-53);** 79772 (1e-36)	**190703 (9e-19)**			**CMJ105C (7e-07);** CMC181C (0.069);	YP_721388 (5e-11) *T.erythraeum*
Tic55	CAA04157	**At2g24820 (Tic55);** *At3g44880 (PAO/ACD1); At4g25650 (PTC52); At1g44446 (CAO)*	**104363 (1e-157); 235252 (1e-153);** *215096 (PAO-1; 6e-47); 158899 (PTC52; 6e-46); 63448 (PAO-2; 3e-42); 108111(CAO-1; 2e-22); 108596 (CAO-2; 2e-22)*	195950 (2e-28); 196533 (3e-28); 196534 (5e-27); 206616 (1e-15); *206612 (5e-19);* *113560 (PAO; 3e-35);* *195951 (CAO; 3e-28);* *114479 (CAO-like; 4e-21)*	26228 (1e-26); 18431 (4e-23); 43309 (5e-52); *35581 (CAO; 2e-15);*	35257 (3e-25); 37319 (1e-24); 24496 (2e-37); *33877 (4e-15)*	*CMG162C (5e-04)*	ZP_00107775 (1e-65) *N. punctiforme*
Tic62	CAC87810	**At3g18890 (Tic62);** At2g34460; At4g31530; At5g02240; At2g37660	*209107 (8e-84); 209343 (1e-28); 109526 (4e-16); 108314 (3e-12); 161093 (9e-11); 110108 (5e-05);*	*142153 (2e-27); 206614 (1e-32); 193756 (4e-25); 205577 (CGLO; 3e-22); 206615 (5e-08); 196378 (6e-06)*	*1559 (2e-20); 17427 (2e-11); 42412(5e-09); 7116(4e-06)*	*19959 (8e-11); 34533 (4e-02)*	*CMM310C (6e-31);*	*NP_486791 (9e-34) N. PCC7120*
Tic32	AAS38575	**At4g23430 (Tic32);** At4g23420; At4g11410; At5g02540; At2g37540	124527 (4e-80); 193394 (1e-78); 151461 (1e-76); 80759 (5e-62); 99498 (9e-64); 218730 (1e-45); 173522 (4e-39); 64665 (7e-48)	206617 (1e-36); *191453 (1e-30);* *145585 (8e-27);*	*38397 (1e-41);* *48682 (9e-36);* *13992 (1e-28);* *38463 (1e-22);* *36833 (9e-28)*	*11084 (1e-37);* *31338 (1e-36)*	CMA135C (5e-19); CMO249C (2e-19); CME017C (4e-18)	NP_485762 (1e-48) *N. PCC7120*
Tic40	CAB50925	**At5g16620 (Tic40)**	**166296 (2e-71); 167840 (8e-63)**	**183027 (8e-39)**	**26145 (2e-38)**	**35165† (8e-40)**		*YP_319945 (0.65) A. variabilis*
Toc64	AAF62870	**At3g17970 (Toc64-III);** At5g09420 (OM64); *At1g08980 (Ami1)*	**107505 (Toc64-1; 1e-159) 105620)Toc64-2; 1e-157) 197643 (1e-143)**	*205899 (Ami2; 3e-32)*	**17396 (5e-64)**	**12917 (5e-77)**		*YP_723678 (2e-22) T. erythraeum (Amidase only)*

The presence (green) or absence (red) of Toc and Tic components within the surveyed genomes is shown. Accession (or protein ID) numbers are recorded, with the *e*-values of the reciprocal BlastP against the *A. thaliana* database in parentheses. A yellow box highlights *C. reinhardtii* results. The most likely ortholog is highlighted in boldface type. Several components are represented by multigene families, whose members are not necessarily functionally equivalent. These homologs are indicated by italics. Asterisks (*) represent genes with incomplete gene models. The dagger (†) indicates a modified gene model.

C-terminal bacterial surface-antigen domain encoding the pore-forming transmembrane β-barrel fold and two N-terminal polypeptide-translocation-associated (POTRA) domains (supplemental Table 1 and Gentle *et al.* 2005). The POTRA domains may facilitate transit peptide interactions (Ertel *et al.* 2005) or protein insertion into the outer membrane (Sánchez-Pulido *et al.* 2003). Both the CrToc75 POTRA motifs are divergent from

their respective canonical PFAM sequences, which is similar to other Toc75 proteins (supplemental Table 1).

CrToc75 is probably targeted to the chloroplast by an N-terminal bipartite targeting motif, similar to PsToc75 (Tranel and Keegstra 1996). PsToc75 encodes a chloroplast transit peptide followed by a polyglycine rich region, both of which are necessary and sufficient for chloroplast outer membrane localization, although the precise function of the polyglycine region remains unclear (Baldwin and Inoue 2006). CrToc75 contains a TargetP-predicted N-terminal transit peptide, albeit one predicted to target to the mitochondria (supplemental Table 1). Nevertheless, the TargetP mitochondrial-targeting prediction is dubious since neural network predictions of *C. reinhardtii* chloroplast transit peptides are inaccurate for reasons that still remain unclear (Franzén *et al.* 1990; Patron and Waller 2007). Following the transit peptide, CrToc75 clearly encodes a polyglycine motif (supplemental Table 1). Experimental evidence confirming that CrToc75 is located in the chloroplast outer membrane is now required.

The second putative Toc75 homolog in *C. reinhardtii* (CrOEP80) shares higher similarity with AtOEP80 than with AtToc75 (not shown). As AtOEP80 has no association with other protein translocation components (Eckart *et al.* 2002), by extension we predict that CrOEP80 is not involved in chloroplast protein import in *C. reinhardtii*. However, CrOEP80 also contains an N-terminal polyglycine motif (supplemental Table 1), which is unexpected since AtOEP80 lacks this targeting motif. The functional significance of this polyglycine region in CrOEP80 is unclear.

Similar to *C. reinhardtii*, the sister green-algal taxa *Ostreococcus* spp. also encode two Toc75 homologs (Table 1). A similar dichotomy is also observed in these genomes, where one putative Toc75 homolog appears orthologous to AtToc75-III, while the other homolog is likely to be orthologous to AtOEP80. Unlike *C. reinhardtii*, however, the bryophyte *P. patens* contains four proteins that appear orthologous to AtToc75-III and one to AtOEP80 (Table 1), consistent with *P. patens* EST data (Hofmann and Theg 2003; Inoue and Potter 2004).

Overall, our analyses detected at least two putative Toc75 homologs in all the Viridiplantae lineages (Figure 1). In the green alga, including *C. reinhardtii*, only one protein is orthologous to AtToc75-III and PsToc75 and hence probably is involved in chloroplast protein translocation. In contrast, *A. thaliana* and *P. patens* encode two and four PsToc75 orthologs, respectively. A recent genome duplication (Rensing *et al.* 2007, 2008) may explain why the haploid *P. patens* genome contains twice the number of PsToc75 orthologs than *A. thaliana*. If so, this suggests that at least two Toc75 paralogs were already present in the green lineage before the divergence of *P. patens*. The increased number of Toc75 homologs may reflect the higher complexity level in multicellular land plants *vs.* single-celled algae.

All the Viridiplantae genomes encode only one ortholog of AtOEP80. Although the function of AtOEP80 is currently unresolved, an attractive hypothesis is that it functions to assemble and insert outer membrane β-barrel proteins, including AtToc75-III (Inoue and Potter 2004). Such a mechanism would be analogous to the function of the Sam50/Tob55 in mitochondria, which facilitates the insertion of Tom40 and other β-barrel proteins (Kozjak *et al.* 2003; Paschen *et al.* 2003).

One putative, but very divergent, Toc75 homolog was identified in *C. merolae* (Table 1). Whether this protein is more similar to AtToc75-III or AtOEP80 is uncertain from the BlastP analysis alone. However, a weakly predicted transit peptide and a short polyglycine motif may function as a two-component leader similar to that of AtToc75-III (supplemental Table 1). Whether this protein is a translocon channel, a membrane insertion factor, or perhaps both, is uncertain.

GTPase receptors Toc34 and Toc159: Toc34 and Toc159 are GTPase proteins that function as chloroplast transit peptide receptors (Bauer *et al.* 2000; Sveshnikova *et al.* 2000). Together with Toc75, both Toc34 and Toc159 constitute the core components of the Toc complex due to their stable interactions with each other and the transit peptide (Waegemann and Soll 1991). Unlike Toc75, neither GTPase protein has a cyanobacterial ortholog, indicating a eukaryotic origin for these receptor components (Reumann *et al.* 2005).

C. reinhardtii encodes only one Toc34 protein (CrToc34) and one Toc159 protein (CrToc159, Table 1). In contrast, *A. thaliana* encodes two paralogs of Toc34 (AtToc33 and AtToc34) and four paralogs of Toc159 (AtToc159, AtToc132, AtToc120, and AtToc90). These numerous *A. thaliana* Toc34 and Toc159 paralogs exhibit distinct expression profiles and form functionally different Toc complexes, allowing the chloroplast to maintain import of nonabundant, nonphotosynthetic proteins while simultaneously importing highly abundant photosynthetic proteins (Bauer *et al.* 2000). Other higher plants are also likely to contain functionally distinct Toc complexes since they encode multiple copies of Toc34 and/or Toc159. For example, spinach and poplar encode at least two Toc34 paralogs and at least three Toc159 paralogs exist in rice (Voigt *et al.* 2005). With only one homolog each of Toc34 and Toc159, protein import into *C. reinhardtii* chloroplasts is unlikely to involve more than one recognition pathway.

CrToc34 contains an N-terminal GTPase domain with a hydrophobic C terminus, similar to other Toc34 homologs (Figure 2, supplemental Table 2). Alignments and motif analysis show that the GTPase of CrToc34 is highly conserved. In particular, critical residues for GTPase dimerization and function, including the D1 dimerization motif and the arginine finger motif (Weibel *et al.* 2003; Yeh *et al.* 2007), are retained in CrToc34 (Figure 2). Similarly, CrToc159 encodes the

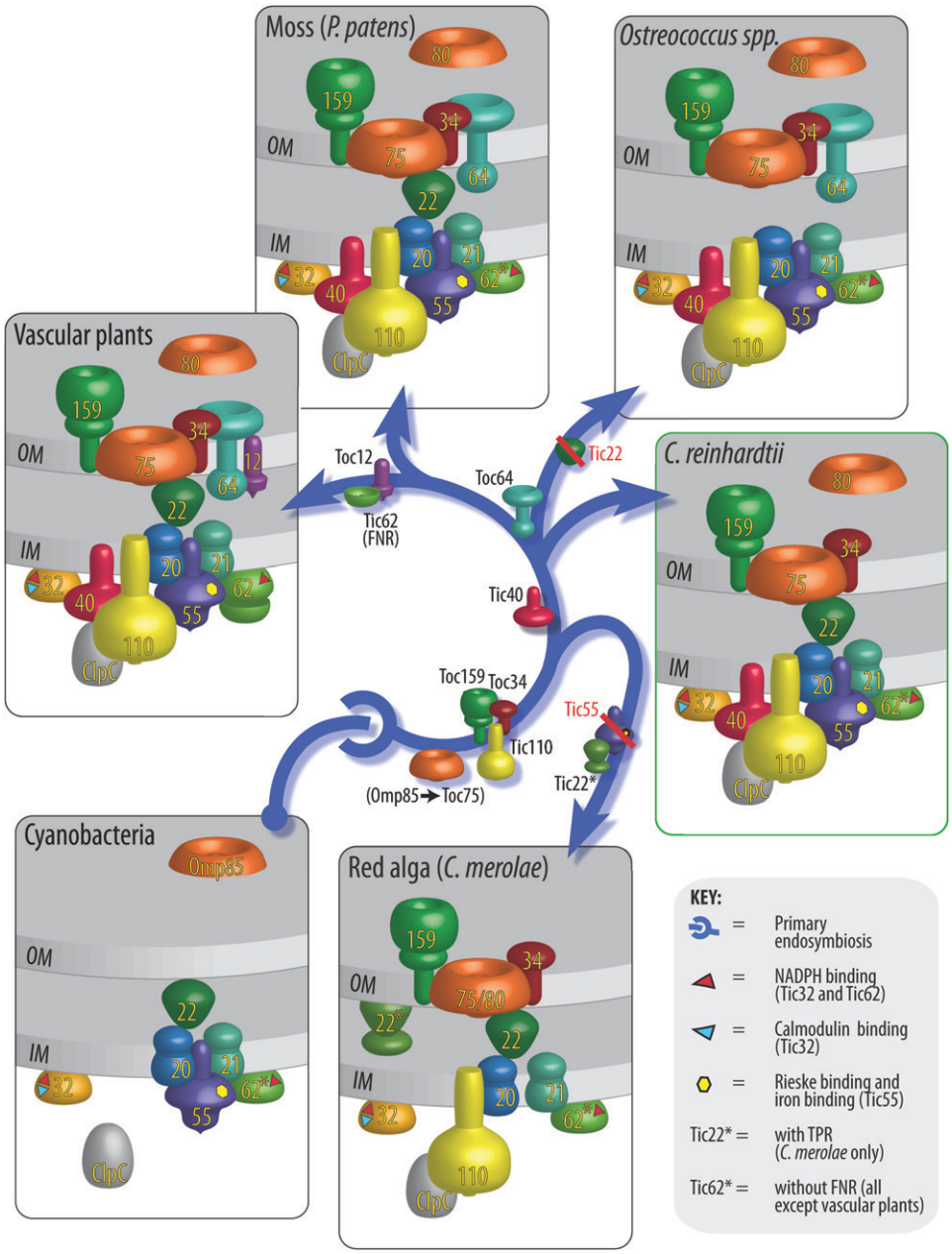

FIGURE 1.—Evolution of the chloroplast protein translocation machinery. Components of Toc and Tic translocons of different phylogenetic lineages are shown, highlighting the overall continuity of the protein import machinery within red algae, chlorophytes, prasinophytes, bryophytes, and vascular plants. First, components derived from the cyanobacterial endosymbiont are also shown, including Omp85, Tic20, Tic22, Tic55, Tic32, and the NAD-binding domain of Tic62, as well as stromal factors such as ClpC. Components acquired early during plastid acquisition are represented in both the red and green lineages, including the conversion of Omp85 to Toc75, Tic110, Toc159, and Toc34. Tic40 was developed after the Viridiplantae diverged, since it is absent from red algae and cyanobacteria. Toc64 is also specific to Viridiplantae, but it is also absent from *C. reinhardtii*. It may have developed after the divergence of the chlorophytes or perhaps was lost specifically from the chlorophyte lineage. Specific to the vascular plant genomes is full-length Tic62, encoding a C-terminal FNR-binding domain. However, the N-terminal NAD-binding domain is found in all surveyed genomes, including cyanobacteria. Finally, Toc12 has been identified only in *P. sativum* to date.

conserved GTPase and hydrophobic domains distinctive of Toc159 proteins (supplemental Table 2). Overall sequence analyses indicate that CrToc159 shares more similarity to AtToc132 than other *A. thaliana* Toc159 paralogs (Table 1).

Unlike higher plants, however, CrToc34 has a significantly longer and more acidic N terminus. Specifically, CrToc34 contains 93 amino acids upstream of the GTPase domain, encoding a total of 21 glutamic acid and 18 aspartic acid residues, or 1 acidic residue for every 2.4 amino acids (supplemental Table 2, Figure 2). In contrast, the N termini of AtToc33 and AtToc34 are less than half the length of CrToc34 and contain only 1 acidic residue per 7.2 amino acids. Although such a highly acidic N-terminal domain in CrToc34 is abnormal for Toc34 homologs, it is strikingly similar to the N-terminal domains of Toc159 homologs, which are defined by their bias for acidic amino acids (BAUER *et al.* 2000).

In contrast, the putative CrToc159 homolog lacks the acidic N-terminal domain distinctive of PsToc159 and three of the four *A. thaliana* Toc159 paralogs (supplemental Table 2). Whereas PsToc159 and AtToc159 contain one acidic residue per 3.7 and 4.0 amino acids, respectively, CrToc159 contains approximately half that number, with one acidic residue every 8.4 amino acids. Thus, in *C. reinhardtii*, the highly acidic transit peptide receptor of the Toc complex is found on CrToc34, not CrToc159, as would have been anticipated from higher plant models. It is currently unclear whether this interchange has an impact on the recognition of

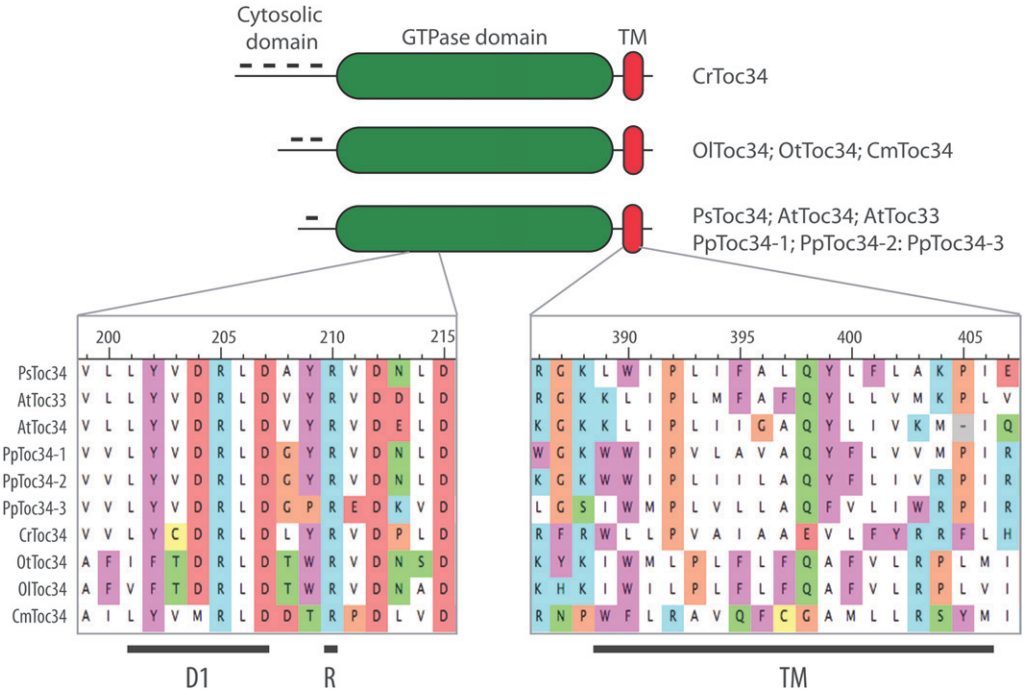

FIGURE 2.—CrToc34 has a negatively charged N terminus. Toc34 alignment, indicating the dimerization motif (D1), arginine finger (R), and the predicted transmembrane helix (TM). The model shows the GTPase (green) and transmembrane helix (red) and highlights the length of and negative charges of the N terminus of CrToc34 (represented by "-").

C. reinhardtii transit peptides. This may be a possibility since the current model for chloroplast import predicts that transit peptides are recognized initially by Toc34 and interact subsequently with Toc159 (BECKER *et al.* 2004b; SOLL and SCHLEIFF 2004). Perhaps the acidic CrToc34 domain influences the composition of *C. reinhardtii* transit peptides to resemble vascular plant mitochondrial transit peptides, which are generally shorter (ZHANG and GLASER 2002) and more enriched in arginine residues within their N termini (PUJOL *et al.* 2007).

The prasinophyte green algae, *O. lucimarinus* and *O. tauri*, also contain single homologs for Toc34 and Toc159. The N-terminal regions of OlToc34 and OtToc34 are slightly longer and more acidic than the higher plant homologs, but are not as acidic as CrToc34 (supplemental Table 2, Figure 2). Overall, global sequence alignments and BlastP analyses of Toc34 proteins indicate that both OlToc34 and OtToc34 are more similar to the Toc34 homologs from *A. thaliana* and *P. sativum* than CrToc34 (Table 1). Correspondingly, the Toc159 homologs, OlToc159 and OtToc159, are more similar to vascular plants than *C. reinhardtii*, containing long and acidic N-terminal domains (supplemental Table 2).

Unlike the green algae, the bryophyte *P. patens* encodes both GTPase receptors in small multigene families, with three putative Toc34 homologs and four predicted Toc159 homologs (Table 1). All the *P. patens* Toc34 proteins have short N termini, which are not significantly enriched with acidic residues (supplemental Table 2), making them more similar to higher plant Toc34 proteins than the green algal orthologs.

All four putative *P. patens* Toc159 homologs share high sequence similarity to AtToc132 (Table 1) and are represented by EST data (supplemental Table 2). HOFMANN and THEG (2003) have previously identified PpToc125 (protein 216964, Table 1). The *P. patens* Toc159 homologs form subgroups consisting of short or long N-terminal domains, arranged in two clusters in a head-to-tail configuration, suggesting two relatively recent gene duplication events (supplemental Figure 2).

Finally, the red alga *C. merolae* also encodes two putative GTPase receptor proteins. On the basis of the size and overall protein structure of these proteins, we assigned one of these proteins as the most likely Toc34 homolog (CMP284C) and the second protein as Toc159-like (CMQ137C). The CmToc34 homolog is the shorter GTPase and shares higher sequence similarity to AtToc34 than the putative CmToc159-like protein. In contrast, the CmToc159-like shares similar features with other Toc159 proteins, including a long N-terminal domain, followed by a GTPase domain and a hydrophobic C terminus (supplemental Table 2). However, CmToc159-like shares only low sequence similarity to AtToc132 (Table 1), its GTPase domain is poorly conserved, and its N-terminal domain is not highly acidic (supplemental Table 2). This poor sequence similarity to other Toc159 proteins may explain why CmToc159-like candidates were overlooked in previous analyses (MCFADDEN and VAN DOOREN 2004).

Our analysis of these numerous genomes indicates that both Toc34 and Toc159 are highly conserved components of the plastid protein import complex, since at least one copy of each GTPase was already present in the common ancestor of red and green algae (Figure 1). This suggests that the heterodimerization of GTPase proteins is likely to be a fundamental process required for chloroplast protein import (SCHLEIFF *et al.*

2003). Also apparent from our analysis is that one pair of GTPase proteins is sufficient for all chloroplast protein import in single-celled green and red algae, whereas multicellular vascular and nonvascular plants utilize multiple import pathways and correspondingly encode more than one pair of GTPase proteins.

A recent study using hydrophobic cluster analysis on AtToc159 implies that the acidic N-terminal and hydrophobic C-terminal domains are derived from ancient duplications of the GTPase domain and that the N-terminal domain of higher plant Toc159 proteins (*e.g.*, *A. thaliana* and *P. patens*) have become more acidic over time (Hernández Torres *et al.* 2007). Hence, the low acidity in the N termini of *C. reinhardtii* and *C. merolae*, and to a lesser extent *Ostreococcus* spp., may represent a relatively more primitive state.

Tic110, Tic20, and Tic21, putative inner-membrane channel components: Whereas Toc75 has always been the best candidate for the central pore-forming component of the outer membrane (Schnell *et al.* 1994), the identity of the inner-membrane translocation channel has been controversial, due in no small part to the experimental inaccessibility of this membrane. Candidates have included Tic110 (Heins *et al.* 2002), Tic20 (Chen *et al.* 2002), and, recently, Tic21 (Teng *et al.* 2006).

Only one Tic110 homolog was found in *C. reinhardtii* (CrTic110), similar to the other genomes surveyed with the exception of *P. patens* (Table 1), which is consistent with results of EST studies (Dávila-Aponte *et al.* 2003; Inaba *et al.* 2005). *P. patens* encodes two Tic110 paralogs, possibly due to a recent genomewide duplication, discussed above. Tic110 is absent from cyanobacterial genomes, confirming that this component has eukaryotic origins (Figure 1; Reumann *et al.* 1999).

The solubility and structure of Tic110 has been the subject of considerable debate. Heins *et al.* (2002) reported that PsTic110 forms a voltage-gated anion channel in the presence of proteoliposomes and presented neural network predictions for a β-barrel transmembrane structure (Heins *et al.* 2002). However, Inaba *et al.* (2003) used circular dichroism spectra of AtTic110 truncations to deduce that the large C-terminal domain is predominantly soluble and α-helical and that a hydrophobic N-terminal region anchors the protein into the inner membrane facing the stroma (Inaba *et al.* 2003). Our sequence analysis is more consistent with this latter model since CrTic110 contains predicted N-terminal transmembrane α-helices and is unlikely to contain a transmembrane β-barrel domain. The recent hidden Markov model prediction algorithm ProfTMB (Bigelow *et al.* 2004) estimates that CrTic110 has only a 20.7% likelihood of forming a transmembrane β-barrel domain (supplemental Table 3). By contrast, CrToc75 has a predicted 72.3% chance of forming a transmembrane β-barrel.

Alignments of Tic110 orthologs show that CrTic110 and its *C. merolae* ortholog are very divergent from higher plant Tic110 homologs (not shown), reflected in the comparatively low BlastP scores (Table 1). Nevertheless, several conserved motifs are revealed by global protein alignments, particularly a leucine-zipper-like motif (L-x-x-L-x-x-x-L-G, supplemental Figure 1) within the C-terminal half of the soluble domain. The functional significance of this motif is not clear, although a truncation mutant of AtTic110 lacking the C-terminal half of the soluble domain has been shown to be defective in chloroplast protein import and exhibits lower binding affinity for the Toc complex (Inaba *et al.* 2005). Alignments also identify another significant motif (F-L-L-P-W-K/R-R, supplemental Figure 1) within the N-terminal half of the soluble domain that is conserved among only the Viridiplantae Tic110 homologs, including CrTic110. This motif lies in the center of the putative transit peptide-binding domain defined by truncation mutants of AtTic110 (Inaba *et al.* 2003, 2005).

Tic20 is an alternate candidate for the inner-membrane translocation channel. PsTic20 is an integral membrane protein with four α-helical transmembrane domains (Chen *et al.* 2002, supplemental Table 4) that interacts with PsTic22 *in vivo* (Kouranov and Schnell 1997). Phylogenetic analyses show that Tic20 is distantly related to other transporters, including cyanobacterial branched-chain amino acid transporters and putative mitochondrial channel proteins, Tim17 and Tim23 (Reumann *et al.* 2005). Significantly, all these proteins also contain four α-helical domains (Mokranjac and Neupert 2005). This phylogenetic and structural relationship to channel-forming proteins strongly implicates Tic20 as a component of the inner-membrane channel, although *in vitro* and *in organellar* evidence is still lacking (Soll and Schleiff 2004).

C. reinhardtii encodes two Tic20 paralogs, but four orthologs were found in *A. thaliana* (Table 1). Since previous studies of *A. thaliana* reported only one (Chen *et al.* 2002) or two (Jackson-Constan and Keegstra 2001) Tic20 homologs, we performed a phylogenetic study of all the Tic20 sequences obtained from the BlastP analyses to determine which was more related to PsTic20. A maximum-likelihood tree demonstrates that the *A. thaliana* homologs form two distinct clades, where AtTic20-I and AtTic20-IV group together with PsTic20, while AtTic20-II and AtTic20-V are more similar to each other than to PsTic20 (Figure 3). On the basis of this phylogeny, only AtTic20-I and AtTic20-IV are likely to be functionally similar to PsTic20, although a role in import cannot be ruled out for the remaining AtTic20 paralogs. Of the two *C. reinhardtii* homologs, one is more similar to AtTic20-I (CrTic20) and is thus likely to be orthologous to PsTic20 (Table 1). Meanwhile, the second homolog is more similar to AtTic20-II and AtTic20-V and may or may not be involved in translocation (CrTic20-like). CrTic20 also comprises four predicted transmembrane α-helices (the first helix is weakly pre-

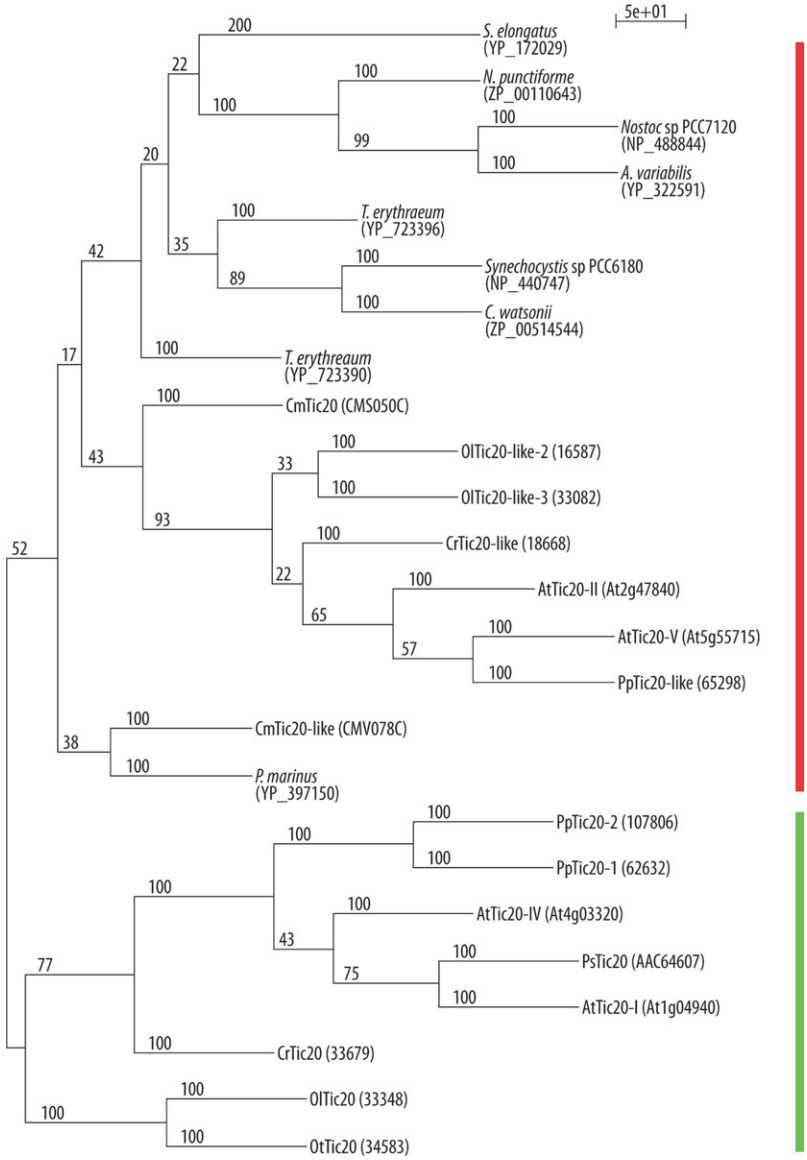

FIGURE 3.—Phylogenetic tree of Tic20. A maximum-likelihood tree drawn from 125 characters with bootstrap values for 100 replicate trees shows that Tic20 proteins form two distinct clades. The first clade (green line) contains PsTic20 orthologs. Each Viridiplantae genome surveyed is represented here. The second clade (red line) contains the cyanobacterial homologs, as well as a subclade that contains Tic20-like proteins.

dicted), as well as a predicted chloroplast-targeted transit peptide (supplemental Table 4).

Similar to *C. reinhardtii*, both prasinophytes *O. lucimarinus* and *O. tauri* contain only one putative ortholog of AtTic20-I. In contrast, the bryophyte *P. patens* genome contains two homologs that share significant similarity to AtTic20-I (Figure 3). However, both PpTic20-1 and PpTic20-2 are more similar to each other and AtTic20-I (and PsTic20) than to AtTic20-IV, indicating that these genes most likely duplicated after the divergence of bryophytes from the higher plant lineage. On the other hand, it is also likely that the duplication event producing AtTic20-IV occurred exclusively within the higher plant lineage, since the green algal and bryophyte genomes lack a detectable ortholog to this protein (Figure 3). The prasinophytes and the bryophyte resemble *C. reinhardtii* in that they encode a single ortholog of AtTic20-II/AtTic20-V (Table 1).

C. merolae also contains two homologs of Tic20, but both proteins appear to be more similar to AtTic20-V than AtTic20-I (Figure 3). Intriguingly, one homolog (CMV078C) is encoded on the chloroplast genome, which is unequivocal proof that Tic20 is endosymbiont derived (REUMANN *et al.* 2005). However, it remains unknown whether the *C. merolae* Tic20 proteins are also channel-forming proteins since neither is clearly orthologous to PsTic20.

Recently, AtTic21 has been proposed as another component of the inner-membrane translocation channel (TENG *et al.* 2006). In contrast, others hypothesize that this protein is a novel chloroplast iron permease, called PIC1 (DUY *et al.* 2007). Our analysis is unable to resolve whether AtTic21/PIC1 is a protein translocator or an iron permease. Signature motifs for permeases and ion channels are identifiable in Tic21/PIC1, but support is weak and inconclusive (supplemental Table

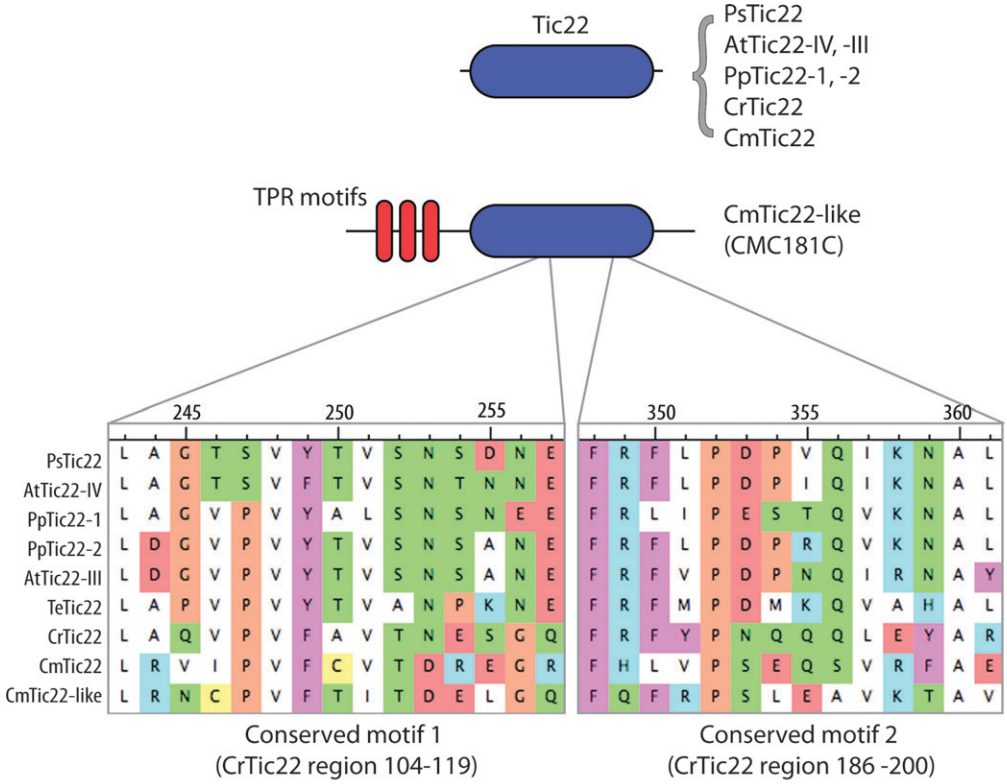

FIGURE 4.—Conserved Tic22 motifs. Alignment of Tic22 proteins highlighting two conserved motifs. The Tic22 domain is represented in blue, while the TPR domains unique to CmTic22-like are shown in red.

5). Our analyses confirm that all Tic21/PIC1 homologs contain four predicted transmembrane α-helices and are highly conserved among photosynthetic organisms (Table 1). Although *A. thaliana* encodes only one Tic21/PIC1, our BlastP approach detected two *C. reinhardtii* paralogs (CrTic21/PIC1-1 and CrTic21/PIC1-2, supplemental Table 5).

Tic22, the intermembrane space adapter: Tic22 is a globular protein associated with the intermembrane space side of the chloroplast inner membrane (KOURANOV *et al.* 1998). This localization, along with its interactions with other Toc and Tic components (KOURANOV *et al.* 1998; HÖRMANN *et al.* 2004), implicates Tic22 as a scaffold protein between the Toc and Tic translocons, which may guide inbound proteins from the outer to the inner membrane (KOURANOV *et al.* 1998; SOLL and SCHLEIFF 2004).

A single putative *C. reinhardtii* Tic22 was identified (CrTic22, Table 1), whereas *A. thaliana* contains two Tic22 paralogs. Both AtTic22 proteins contain predicted chloroplast transit peptides; however, CrTic22 lacks a canonical chloroplast transit peptide (supplemental Table 6). Instead, it encodes a short N-terminal leader of 30 amino acids upstream of the conserved Tic22 domain, which may function as a transit peptide to the intermembrane space. This is consistent with recent studies that indicate that PsTic22 contains an atypical N-terminal leader that interacts with the outer membrane Toc complex, but does not enter the stroma (VOJTA *et al.* 2004). Similarly, the two *P. patens* Tic22 homologs detected with BlastP searches also lack predicted transit peptides, but have long N-terminal extensions (supplemental Table 6).

Unexpectedly, extensive analysis of the prasinophyte *O. lucimarinus* and *O. tauri* genomes did not detect any Tic22 homologs (Table 1). Numerous searches based on several homologs from a wide range of genomes using protein- and gene-based Blast algorithms failed to detect prasinophyte Tic22 orthologs. We also focused the search by using short, conserved regions of Tic22 identified from multiple sequence alignments (Figure 4), to no avail. Finally, Tic22 is not encoded on the recently published *O. tauri* chloroplast or mitochondrial genomes (ROBBENS *et al.* 2007). Given the general conservation of Tic22, it seems highly unlikely that this gene has been overlooked in *Ostreococcus* spp., especially considering that these genomes are small and very compact (DERELLE *et al.* 2006; PALENIK *et al.* 2007) and that Tic22 is absent from both species (Table 1). The apparent lack of Tic22 in these two prasinophytes, but its presence in *C. reinhardtii*, suggests a specific loss in the prasinophyte lineage after the divergence of chlorophytes. Absence of Tic22 in prasinophytes begs the question of just how essential this otherwise ubiquitous protein is for plastid protein import.

One putative nuclear-encoded Tic22 ortholog was identified in *C. merolae* (Table 1), but none within its chloroplast genome. This contradicts a previous study that reported a second Tic22 homolog encoded on the chloroplast genome itself. However, a second Tic22-

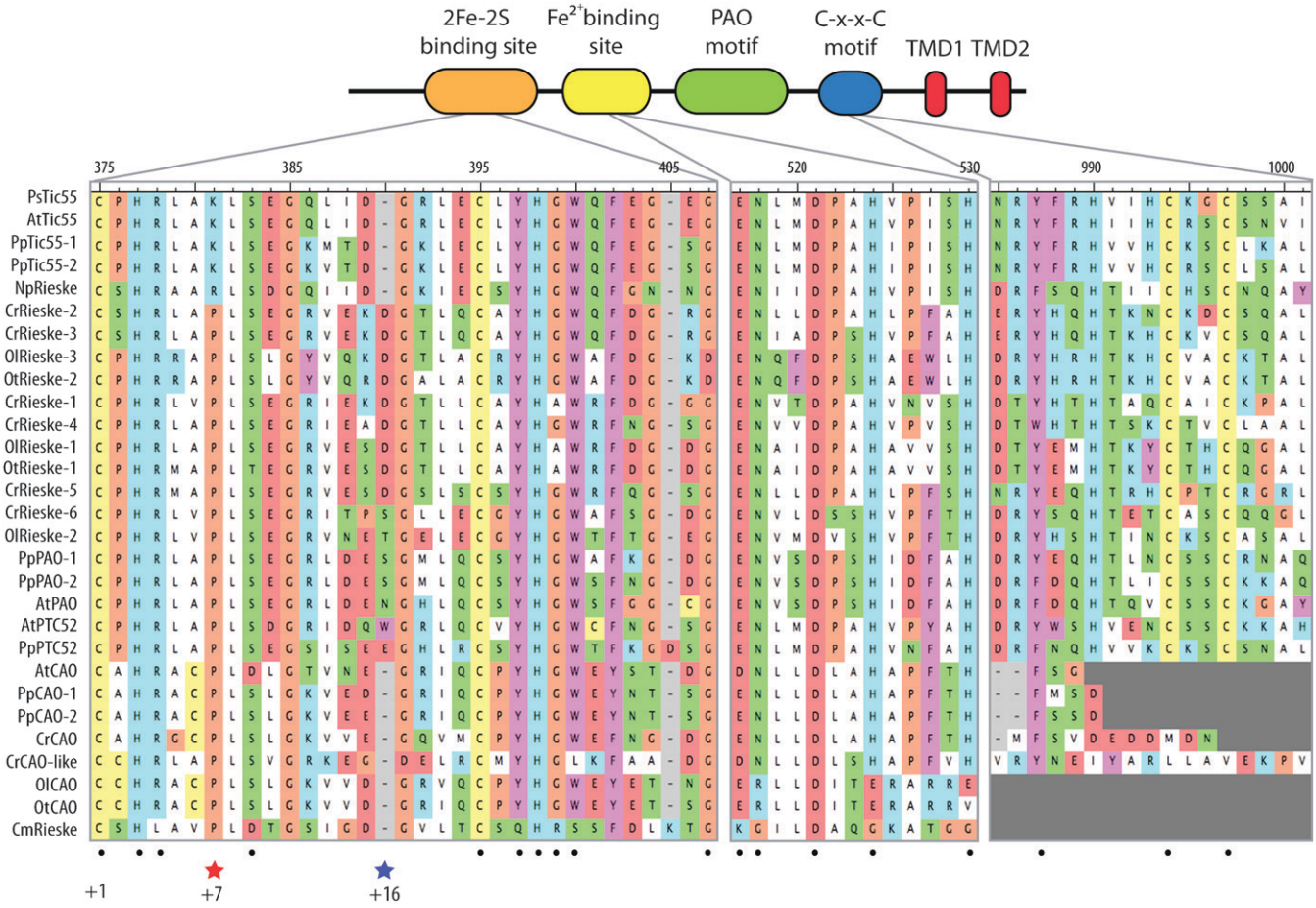

FIGURE 5.—Tic55 protein structure. Tic55 overview including the Rieske motif (orange), the mononuclear iron-binding site (yellow), the C-x-x-C motif (blue), two transmembrane domains (red), and the PFAM PAO motif (green). The alignment highlights conserved residues within these domains (black dots). Also indicated are positions where Tic55 orthologs encode a basic residue, instead of a proline at position +7 (red star), and an indel that is absent from Tic55 at position +16 (blue star).

domain-bearing protein modified with N-terminal tetratricopeptide repeat (TPR) motifs was detected (CmTic22-like, Figure 4). An ortholog also occurs in the related red alga *Galdieria sulphuraria* (accession no. 07901.g27.t1). The fact that protein–protein interaction domains such as TPR motifs are found on a Tic22 protein suggests that the Tic22 domain itself may also be an interaction element, which supports current models of Tic22 function.

Tic55, Tic62, and Tic32, predicted redox regulators of import: The chloroplast import complex also consists of several proteins peripheral to the translocation channels. Tic55 (CALIEBE *et al.* 1997), Tic62 (KÜCHLER *et al.* 2002), and Tic32 (HÖRMANN *et al.* 2004) are inner-membrane proteins hypothesized to be redox-state receptors, regulators of protein import that respond to the metabolic state of the chloroplast.

Tic55 is an integral inner-membrane protein that contains a stromal-localized Rieske-type iron sulfur (2Fe-2S) cluster, as well as a mononuclear iron (Fe^{2+})-binding site. Initially identified as a constituent of the *P. sativum* Tic translocon (CALIEBE *et al.* 1997), Tic55 is phylogenetically related to three other non-heme iron-binding proteins: chlorophyllide *a* oxygenase (CAO), protochlorophyllide oxidoreductase *a* translocation complex (Ptc52), and pheophorbide *a* oxygenase (PAO) (GRAY *et al.* 2004). Since all these closely related proteins have roles in chlorophyll metabolism, GRAY *et al.* (2004) hypothesize that Tic55 is also involved in the metabolism of chlorophyll, which may not be mutually exclusive to its putative role in chloroplast import.

C. reinhardtii contains eight Rieske-type non-heme iron-binding oxygenases (CrTic55-like proteins, Table 1), although sequence analysis alone is insufficient to determine if the chlorophyte encodes a functional ortholog of PsTic55 (Figure 5 and supplemental Figure 3). Four of the CrTic55-like proteins lack domains distinctive of Tic55 orthologs, such as C-terminal transmembrane helices or the conserved C-x-x-C motif (supplemental Figure 3, GRAY *et al.* 2004). However, further scrutiny of the four remaining proteins casts doubt on whether they are Tic55 orthologs because they still lack key residues characteristic of other Tic55 proteins, from cyanobacteria to vascular plants (Figure 5). Specifically,

these *C. reinhardtii* Tic55-like sequences (CrTic55-like 1-4) encode a proline instead of a basic amino acid at position +7 and lack a deletion at position +16 of the Rieske 2Fe-2S motif, which makes them more similar to AtPtc52 than to AtTic55 (Figure 5, Gray *et al.* 2004). This classification is also supported by motif-based homology searches (supplemental Table 7). Further work is required to determine if any of the *C. reinhardtii* Tic55-like proteins is a functional ortholog of Tic55.

Similar analyses of *O. lucimarinus* and *O. tauri* also failed to identify a distinctive Tic55 ortholog from among four candidates in each genome (Table 1). As for *C. reinhardtii*, the potential Tic55 orthologs in the prasinophytes are more similar to AtPtc52 (Figure 5 and supplemental Figure 3). Presumably the cyanobacterial-derived Tic55 homolog was significantly modified or completely lost in the common ancestor of prasinophytes (*Ostreococcus* spp.) and chlorophytes (*C. reinhardtii*), before they diverged but after the divergence of land plants.

In contrast to the green algae, two *P. patens* proteins (PpTic55-1 and PpTic55-2) appear orthologous to AtTic55 (Figure 5 and supplemental Figure 3). Both putative orthologs share high sequence similarity to each other and to AtTic55 (Table 1) and possibly derive from the recent genome duplication discussed previously. In addition, orthologs of AtPAO, AtPtc52, and AtCAO are clearly distinguishable in the *P. patens* genome on the basis of the criteria discussed above (supplemental Figure 3).

Unexpectedly, the red alga *C. merolae* appears to lack orthologs for any of the Tic55 family of proteins (Table 1). Only one Rieske-type protein was detected in our BlastP analysis (CmRieske), but it is unrelated to Tic55 since it lacks the Fe^{2+}-binding and the PAO domains and other Tic55 signature motifs (supplemental Figure 3). Searches of other red algal EST databases and the Galdiera genome also failed to detect homologs of Tic55. The absence of Tic55 from red algae also indicates that it is not essential for chloroplast import, corroborating the apparent absence of distinctive orthologs in *C. reinhardtii* and *Ostreococcus* spp.

Tic62 (Küchler *et al.* 2002) and Tic32 (Hörmann *et al.* 2004) are also peripheral translocation components with predicted roles in redox regulation of chloroplast import. Both proteins bind the coenzyme NADPH and belong to the short-chain dehydrogenase (SDR) superfamily, although sequence analysis reveals that Tic62 belongs to the "extended" family while Tic32 belongs to the "classical" family (Figures 6 and 7, Kallberg *et al.* 2002). Recent *in vitro* data show that both Tic62 and Tic32 act as redox-sensitive molecular switches, since neither protein interacts with Tic110 in the presence of NADPH (Chigri *et al.* 2006).

C. reinhardtii does not encode an ortholog of the full-length PsTic62 or AtTic62, which are bimodular proteins consisting of an N-terminal coenzyme-binding domain and a C-terminal ferredoxin-NADPH-reductase (FNR)-binding domain (Figure 6). Previous studies also found that these bimodular proteins have been observed only in vascular plants (Küchler *et al.* 2002; Balsera *et al.* 2007). However, there are six *C. reinhardtii* proteins homologous to the N-terminal domain of PsTic62 alone (Table 1). We classified these proteins as homologs since they contain a G-x-x-G-x-x-G motif within their coenzyme-binding domains (Figure 6), a characteristic of the Tic62 family of extended SDR proteins (Balsera *et al.* 2007). The *A. thaliana* genome also encodes paralogs of Tic62 that lack the FNR binding (Table 1, Balsera *et al.* 2007), but none of these shorter proteins is known to interact with the translocation complex. How the absence of the FNR domain impacts on the function the CrTic62-NAD proteins is unknown, and it remains unresolved whether

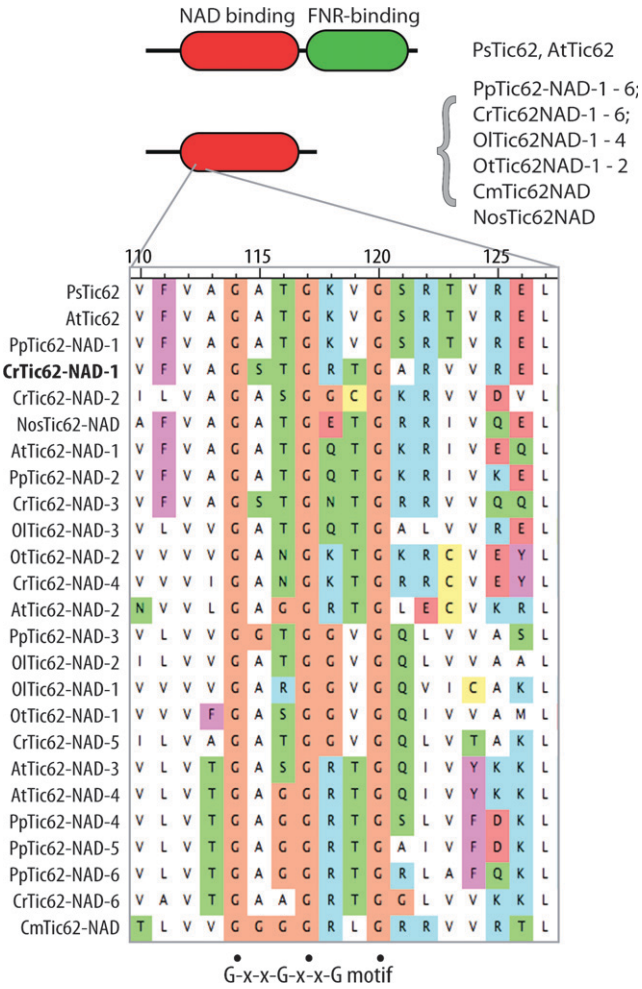

Figure 6.—Tic62 is an extended short-chain dehydrogenase. The conserved domains of full-length Tic62 are modeled with an N-terminal NAD-binding domain, which is classified within the "extended" class of short-chain dehydrogenases (red), and a C-terminal FNR-binding domain (green). Homologs from nonvascular genomes encode only the NAD-binding domain. The signature motif of extended SDR proteins is shown (G-x-x-G-x-x-G, black dots).

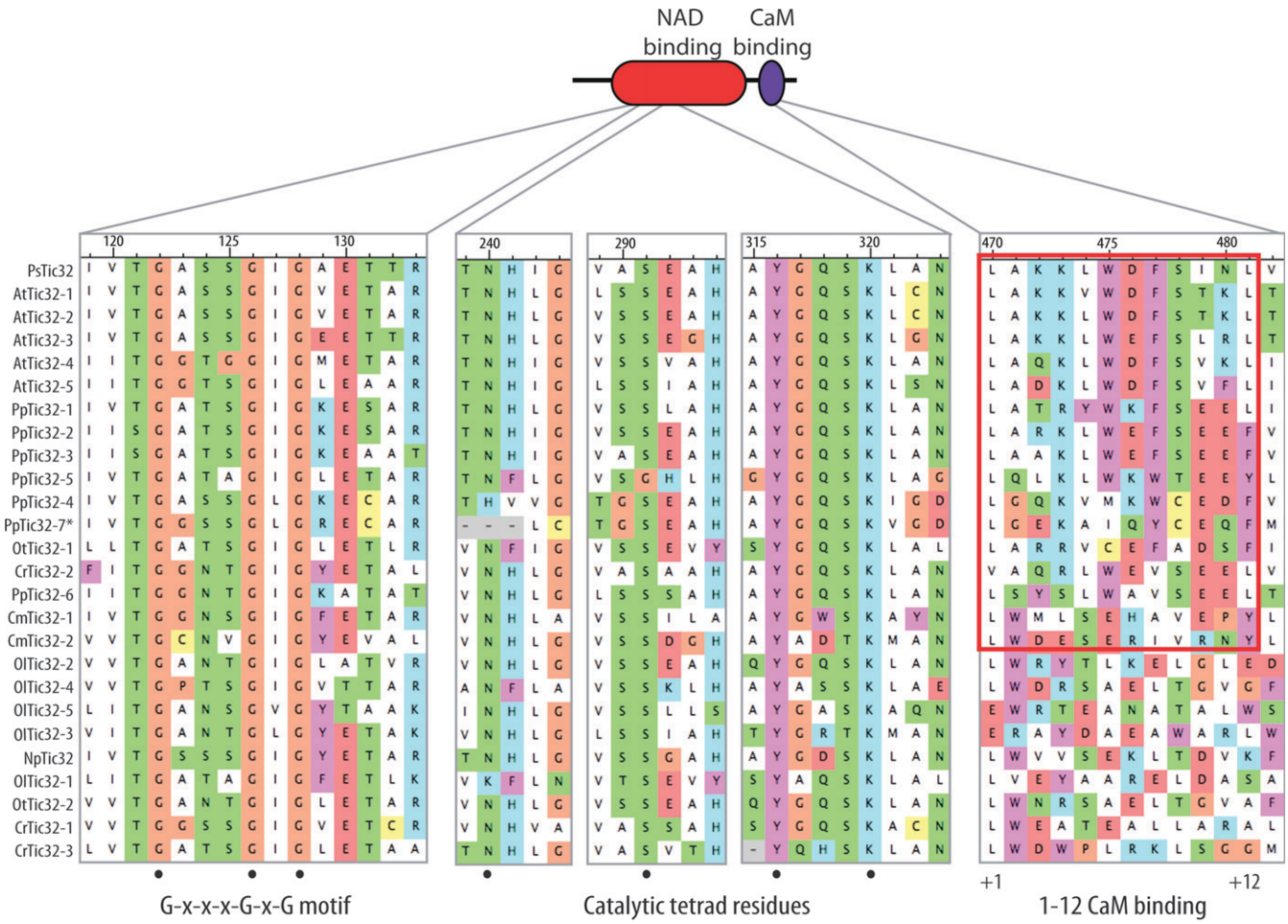

FIGURE 7.—Tic32 is a classical short-chain dehydrogenase. Tic32 domains show the NAD-binding domain (red) and C-terminal 1-12 CaM-binding motif (purple). Indicated are the signature motif of classical SDR proteins (G-x-G-x-x-x-G) and catalytic tetrad residues (black dots). The majority of Viridiplantae sequences encode these residues in a Y-G-Q-S-K motif. A red box indicates the 1-12 CaM-binding motif, defined as bulky hydrophobic residues at positions +1 and +12.

any of these six homologs is actually a component of the import machinery.

The *Ostreococcus* spp., *P. patens*, and *C. merolae* genomes also lack a bimodular Tic62 consisting of both the coenzyme-binding and FNR-binding domains. Our BlastP analysis also detected extended SDR enzymes within these lineages; however, it is not possible to determine if any are functionally orthologous to Tic62 from sequence analysis alone.

Whereas full-length PsTic62 binds both NADPH and FNR in vascular plants, Tic32 binds NADPH and Ca^{2+} via a C-terminal calmodulin-binding domain (CHIGRI *et al.* 2006). Thus chloroplast protein import is sensitive to intracellular Ca^{2+} signaling since the calmodulin-binding domain modulates NADPH binding, which in turn regulates the interactions between Tic32 and Tic110 (CHIGRI *et al.* 2006). A knockout of one *A. thaliana* Tic32 homolog (At4g23430) is embryonic lethal, indicating an essential role in chloroplast biogenesis (HÖRMANN *et al.* 2004).

Three putative *C. reinhardtii* Tic32 sequences were identified in *C. reinhardtii* (CrTic32-like proteins, Table 1). Similar to Tic62, sequence analysis alone is insufficient to definitively label these CrTic32-like proteins as functional orthologs of Tic32. Although the coenzyme-binding domain is highly conserved, particularly within the T-G-x-x-x-G-x-G motif that classifies them as classical SDR enzymes (OPPERMANN *et al.* 2003), the C-terminal domain is not conserved (Figure 7). Consequently, the loosely defined "1-12" calmodulin-binding domain, distinctive of Tic32 orthologs (CHIGRI *et al.* 2006), may not be present. Since classical SDR domains are highly conserved and found in a very large family of enzymes with a wide range of functions (OPPERMANN *et al.* 2003), the detection of a classical SDR domain does not necessarily imply Tic32 function. In particular, Blast analysis against the NCBI database indicates that CrTic32-like-2 and CrTic32-like-3 are more similar to retinol dehydrogenases than to AtTic32 (not shown). Further, CrTic32-like-2 contains a putative mitochondrial transit peptide, making it an unlikely ortholog of Tic32 since both PsTic32 and AtTic32 lack any organellar targeting peptide (HÖRMANN *et al.* 2004). On the other hand, reciprocal BlastP analysis of CrTic32-like-1 does detect

AtTic32 and PsTic32 as best-hit results (Table 1), making it the best candidate for a Tic32 ortholog, despite lacking a calmodulin-binding domain (Figure 1).

We found two homologs of Tic32 in *O. tauri* and five in *O. lucimarinus* (Table 1). As for the *C. reinhardtii* homologs, the N-terminal coenzyme-binding domain is well conserved, but the C-terminal region is not (Figure 7). Furthermore, all the prasinophyte Tic32 homologs have predicted N-terminal-targeting peptides to either the chloroplast or the mitochondria (supplemental Table 9). This also suggests that these proteins are classical SDRs with no function in chloroplast import.

In contrast to the green algae, *P. patens* encodes seven putative homologs of Tic32 (Table 1), all of which contain predicted C-terminal 1-12 calmodulin-binding domains and lack predicted transit peptides (Figure 7, supplemental Table 9), an analogous situation to the *A. thaliana* Tic32 paralogs. However, two are likely to be nonfunctional since they lack a complete catalytic tetrad (PpTic32-4 and PpTic32-5).

Finally, *C. merolae* encodes two classical SDR proteins with C-terminal calmodulin-binding domains (Table 1, Figure 7), although both have different catalytic tetrad motifs (Figure 7) and at least one predicted N-terminal transmembrane helix and are predicted to target to the mitochondria (supplemental Table 9). These features make them unlikely Tic32 orthologs; however, further experimental characterization is required.

Tic40, Toc64, and Toc12, predicted chaperone-binding proteins: Several peripheral Toc and Tic components, including Tic40 (Stahl *et al.* 1999) and Toc64 (Sohrt and Soll 2000), contain domains distinctive for chaperone binding. These interactions are vital for translocon function since chaperones have several important functions in chloroplast protein import—from forming cytosolic targeting complexes (reviewed in Jarvis and Soll 2002) to providing the motor that drives protein translocation (Akita *et al.* 1997). Toc12 is also likely to interact with a chaperone, since it consists predominantly of a DnaJ (Hsp40) domain, the regulatory partner of Hsp70. However, this protein has been characterized only in *P. sativum* (Becker *et al.* 2004a) and currently lacks an *A. thaliana* ortholog. Its sequence is too short, and too redundant, to be utilized by Blast analysis alone and is not discussed further here.

Tic40 is an integral membrane protein with an N-terminal α-helical transmembrane domain, followed by a large soluble domain facing the stroma (Stahl *et al.* 1999; Li and Schnell 2006). This stromal domain is composed of a TPR motif that interacts with Tic110 (Chou *et al.* 2003, 2006) and a C-terminal Hip/Hop chaperone interaction domain (Stahl *et al.* 1999) that interacts with stromal ClpC bound to ATP (Chou *et al.* 2006). Double knockout mutants of Tic40 and Tic110, or Tic40 and ClpC, produce no additive phenotype, providing genetic evidence of these interactions (Kovacheva *et al.* 2005).

C. reinhardtii has a single ortholog of Tic40 (CrTic40) that contains the conserved C-terminal TPR motif and the Hip/Hop chaperone interaction domain (supplemental Table 10). Two asparagine residues implicated in stimulating ClpC ATPase activity (Chou *et al.* 2006) are also conserved (Figure 8). A novel N-terminal motif, h-h-W-h-G-h-G-V-G/h-h (where "h" represents a hydrophobic amino acid), is located downstream of the transit peptide of CrTic40 (Figure 8). This motif is highly conserved across all Tic40 homologs, which is surprising, given that it occurs within the predicted α-helical transmembrane domain and presumably lies within the lipid bilayer (supplemental Table 10 and Chou *et al.* 2003). This motif may represent the recognition site for the second processing event, where the intermediate stromal form of Tic40 (without a transit peptide) is inserted into the membrane and processed to become mature Tic40. Location of this site within the membrane supports the hypothesis that the second protease is a membrane protein (Li and Schnell 2006; Tripp *et al.* 2007). The serine–proline-rich flanking region of the N-terminal transmembrane domain required for membrane insertion in AtTic40 (Tripp *et al.* 2007) is conserved in CrTic40, but more enriched in prolines than in serines (supplemental Table 10).

Similar to CrTic40, both *Ostreococcus* spp. genomes contain one ortholog of Tic40 (OlTic40 and OtTic40, Table 1), which also encodes the conserved transmembrane domain recognition motif (Figure 8) and flanking proline-rich regions (supplemental Table 10). Unlike CrTic40, the prasinophyte Tic40 proteins contain aspartic acid residues in key positions of the Hip/Hop chaperone-binding domain, instead of asparagine (Figure 8). This substitution makes OlTic40 and OtTic40 more similar to the Hip/Hop domains of other eukaryotes species (Chou *et al.* 2006). Unlike the green algae, the two *P. patens* Tic40 orthologs (PpTic40-1 and PpTic40-2, Table 1) contain large N-terminal extensions of unknown function and membrane insertion sites depleted in proline residues, but significantly enriched in serine residues (supplemental Table 10).

Tic40 is absent from cyanobacterial genomes and the red alga *C. merolae* (Table 1), as well as *G. sulphuraria* (not shown). This indicates that Tic40 is specific to the Viridiplantae lineages, most likely by an acquisition after the branching of the red algae (Figure 1).

In vascular plants, Toc64 is an integral membrane protein in the outer membrane, consisting of an inactive amidase domain and a C-terminal TPR motif (Sohrt and Soll 2000). However, *C. reinhardtii* does not encode a full-length Toc64 ortholog (Table 1). An amidase protein was identified, but it lacks the characteristic C-terminal TPR motifs, is more similar to AtAmi1, and contains a presumably active catalytic serine residue (Figure 9). The absence of Toc64 from *C. reinhardtii* supports the view that this component is not essential for protein import, which has been reported previously for both *P. patens* (Hofmann and Theg 2003) and *A.*

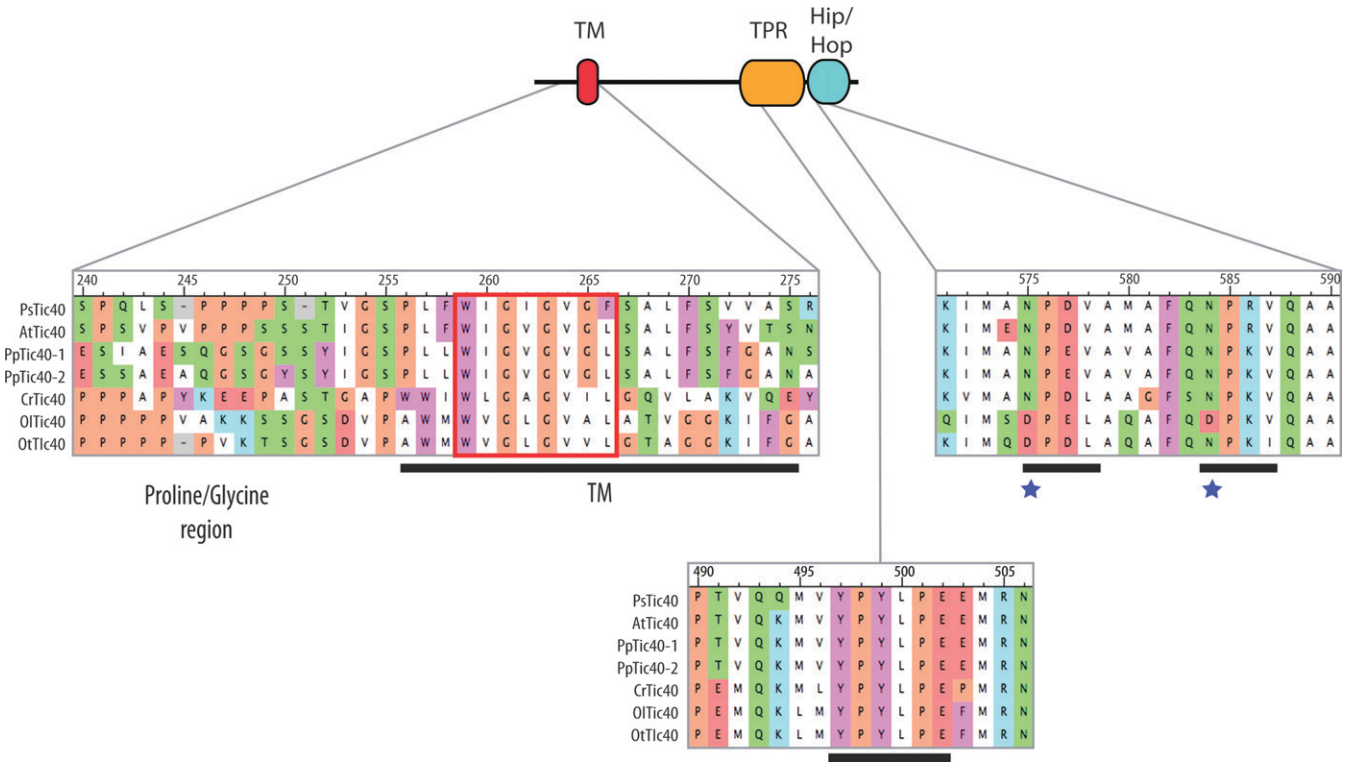

FIGURE 8.—Tic40 conserved motifs. The transmembrane domain (red), TPR motif (orange), and Hip/Hop domain (blue) of Tic40 proteins are shown. The putative processing site (W-h-G-h-G-h) is highlighted in a red box within the transmembrane domain. The conserved Y-P-Y-L-P-E motif within the TPR motif is also shown (black underlines), along with two chaperone-binding sites. Blue stars indicate the conserved asparagine/aspartic acid residues.

thaliana (ARONSSON *et al.* 2007). Further, the absence of Toc64 in *C. reinhardtii* has repercussions for targeting to the mitochondrion, where higher plants utilize a Toc64 homolog (OM64) instead of Tom70, a TPR protein found in other eukaryotes (CHEW *et al.* 2004). Preliminary searches also failed to detect Tom70 in the *C. reinhardti* (not shown).

In contrast, both *Ostreococcus* spp. genomes contain a full-length Toc64 ortholog, including the TPR domain (Table 1). However, the prasinophyte orthologs retain a serine residue in the catalytic triad of the amidase, which distinguishes them from the vascular plant Toc64 proteins (Figure 9) (SOHRT and SOLL 2000; QBADOU *et al.* 2007). Similarly, the three full-length Toc64 orthologs detected in *P. patens* (Table 1) contain the catalytic serine residue within their amidase domains (Figure 9). Two of these PpToc64 orthologs had been identified previously from EST data (HOFMANN and THEG 2003, 2005), and our BlastP approach detected a third ortholog. It is likely that PpToc64-3 is expressed at relatively low levels, since it is represented by fewer ESTs than its paralogs, and a double knockout of PpToc64-1 and PpToc64-2 could not detect any expression of Toc64 in the moss (HOFMANN and THEG 2005). Since the function of the amidase domain in Toc64 remains unresolved, the significance of an active or inactive catalytic domain is also uncertain. However, a requirement of the putative prasinophyte and bryophyte Toc64 homologs for chloroplast protein import cannot be assumed since all the homologs probably contain active amidase domains and may have activities different from the vascular plant proteins (Figure 9).

Cyanobacteria and the red alga *C. merolae* also appear to lack Toc64 homologs (Table 1). Thus Toc64 is likely to be of eukaryotic origin (REUMANN *et al.* 2005), probably appearing after the divergence of red algae (Figure 1). Its presence in both prasinophyte and streptophyte genomes indicates that a full-length Toc64 was present before the divergence of higher plants, but a full-length version was either absent or lost from the ancestor of the chlorophyte *C. reinhardtii*.

Concluding remarks: The focus of chloroplast protein translocation research has progressed from the identification of components to studying how these components interact, how they localize to the chloroplast, and how they function in regulating chloroplast import and biogenesis. The availability of new genomic data from a range of photosynthetic organisms, including *C. reinhardtii*, has the potential to assist with these questions. Our analyses highlight the overall continuity of the chloroplast import machinery, supporting the hypothesis that there was just one primary endosymbiotic event. However, several distinct differences occur between red algae and the Viridiplantae lineages, as well as unanticipated differences within the Viridiplantae (Figure 1).

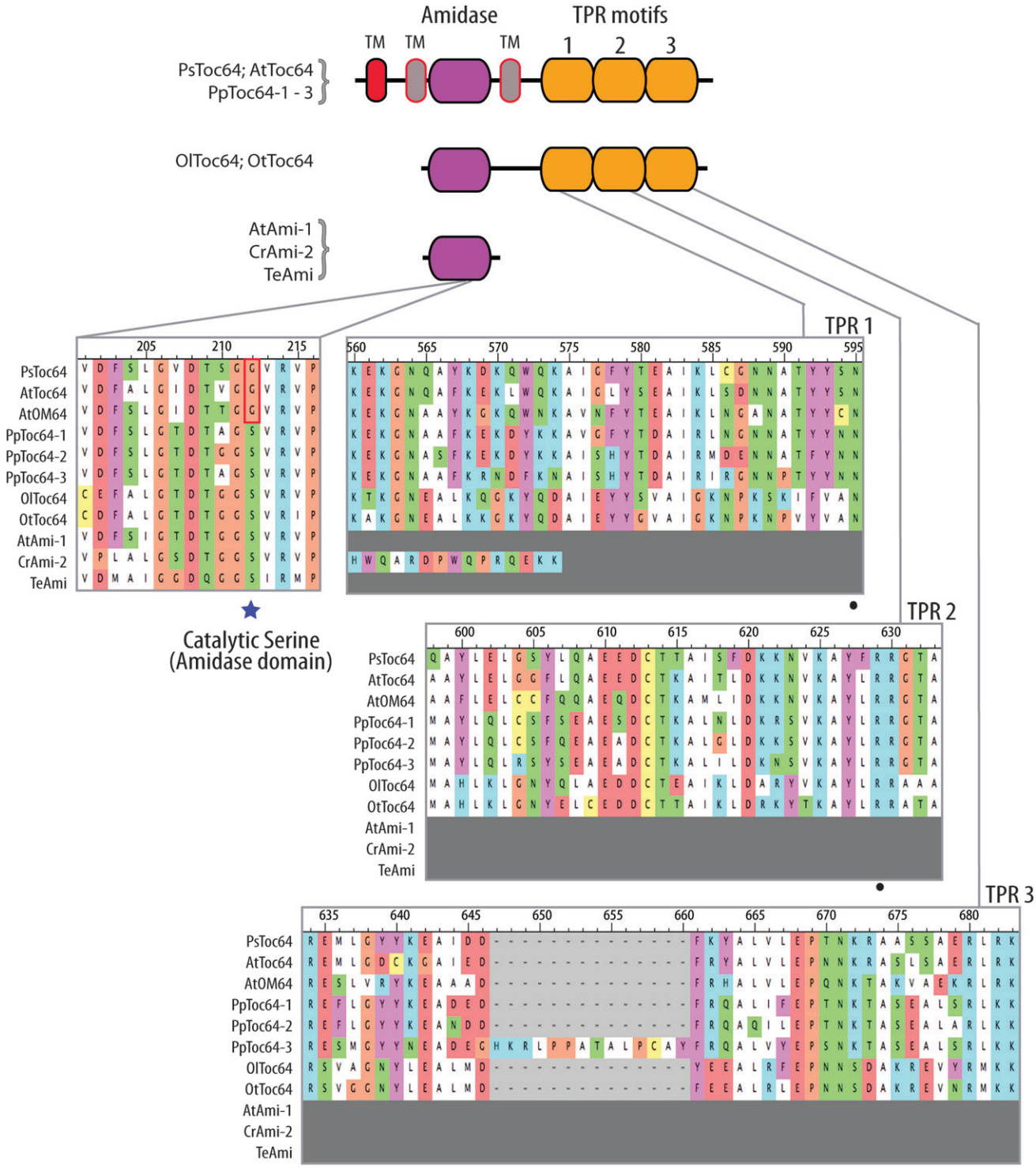

FIGURE 9.—Toc64 conserved motifs. Domains of Toc64 are shown, including the amidase domain (purple), three TPR domains (orange), and transmembrane helices (red). Putative transmembrane domains are indicated by red outlining (QBADOU *et al.* 2007). Highlighted is the catalytic site of the amidase (blue star), inactive glycine residues are shown in a red box and presumably active serine residues below. Alignments show that the TPR motifs are conserved, but are completely absent from CrToc64.

The translocation complexes of *C. reinhardtii* are similar to the model Toc and Tic tranlocons of *A. thaliana*. Conspicuously absent from the chlorophyte, however, are the FNR-binding domain of Tic62, which occurs only in vascular plants, and a full-length Toc64 homolog. The absence of Toc64, and more specifically its cytosolic TPR domain, suggests that Toc64 is not an absolute requirement for protein import and that *C. reinhardtii* lacks the Hsp90-dependent transit peptide chloroplast-trafficking pathway observed in vascular plants (QBADOU *et al.*

2006). Absence of the C-terminal FNR-binding domain of Tic62 in *C. reinhardtii* also indicates that this domain is not essential for protein translocation.

An enduring puzzle of chloroplast protein import in *C. reinhardtii* is the inability of current predictors such as TargetP to reliably discriminate between mitochondrial and chloroplast transit peptides (Franzén *et al.* 1990; Patron and Waller 2007). This analytical limitation suggests that a slightly different transit peptide–receptor recognition mechanism exists in *C. reinhardtii*. Perhaps the molecular reason for this difference is the unusually long and acidic N-terminal cytosolic region of CrToc34. These negatively charged regions on an essential chloroplast import receptor are likely reciprocated by extra positive charges on inbound transit peptides, particularly given that Toc34 is proposed to form the initial contact with the transit peptide (Becker *et al.* 2004b). Since mitochondrial-targeting epitopes are also enriched in arginine residues (Pujol *et al.* 2007), *C. reinhardtii* chloroplast transit peptides might appear more similar to mitochondrial-targeting peptides when screened by neural network predictors. How *C. reinhardtii* distinguishes between chloroplast and mitochondrial proteins is unknown, although a *C. reinhardtii*-specific transit peptide predictor may assist in answering this question.

The *C. reinhardtii* Toc and Tic translocons also resemble the import complexes of the prasinophyte green algae, with two distinct exceptions: the absence of Tic22 and the presence of a Toc64 ortholog containing C-terminal TPR motifs in both *O. lucimarinus* and *O. tauri*. In comparison to vascular plants, *C. reinhardtii* and *Ostreococcus* spp. do not exhibit the trend toward gene multiplication observed in *P. patens* and *A. thaliana*. This trend may reflect an increasing complexity in multicellular plants, where regulation of chloroplast function and biogenesis involves different tissues and developmental stages not required by unicellular systems.

All the Viridiplantae genomes, but not the red algal genome, encode at least two distinct paralogs of Toc75 (Table 1). That *C. merolae* encodes only a single, divergent homolog of Toc75 is surprising. The presence of other Toc components, such as Toc34 and Toc159, as well as the requirement of protein translocation in the initial stages of endosymbiosis, suggests that a functionally equivalent Toc75 channel is present. However, insertion of this protein into the outer membrane presumably requires an Omp85-like activity. The complete genome of the glaucocystophyte *Cyanophora paradoxa* may be instrumental in resolving this issue. The presence of Toc159 and Toc34 in *C. merolae* also suggests that these GTPase receptors were already present in the common ancestor of red algae and the Viridiplantae lineages. Ubiquity of these receptors is indicative of a fundamental requirement of the early stages of chloroplast evolution for an active mechanism based on heterodimerization of GTPase domains for detecting inbound chloroplast proteins (Schleiff *et al.* 2003). Similarly, Tic110 would have been present before the divergence of the red algae from the green lineages, which is consistent with the early need for the recruitment of stromal chaperones to the site of translocation in the progenitor of extant Toc and Tic translocons. Overall, our results show that these new genomes provide a fresh perspective on chloroplast protein import and reduce over-reliance on the vascular plants models.

We thank Ross Waller, Nicola Patron, John West, and Jillian Ackland for discussions on evolution and phylogenetic analyses. M.K. is supported by an Australian Postgraduate Award. G.I.M. is an Australian Research Council Federation Fellow and a Howard Hughes International Research Scholar.

LITERATURE CITED

Akita, M., E. Nielsen and K. Keegstra, 1997 Identification of protein transport complexes in the chloroplastic envelope membranes via chemical cross-linking. J. Cell Biol. **136:** 983–994.

Aronsson, H., P. Boij, R. Patel, A. Wardle, M. Töpel *et al.*, 2007 Toc64/OEP64 is not essential for the efficient import of proteins into chloroplasts in Arabidopsis thaliana. Plant J. **52:** 53–68.

Baldwin, A. J., and K. Inoue, 2006 The most C-terminal tri-glycine segment within the polyglycine stretch of the pea Toc75 transit peptide plays a critical role for targeting the protein to the chloroplast outer envelope membrane. FEBS J. **273:** 1547–1555.

Baldwin, A., A. Wardle, R. Patel, P. Dudley, S. K. Park *et al.*, 2005 A molecular-genetic study of the Arabidopsis Toc75 gene family. Plant Physiol. **138:** 715–733.

Balsera, M., A. Stengel, J. Soll and B. Bölter, 2007 Tic62: a protein family from metabolism to protein translocation. BMC Evol. Biol. **7:** 43.

Bauer, J., K. Chen, A. Hiltbunner, E. Wehrli, M. Eugster *et al.*, 2000 The major protein import receptor of plastids is essential for chloroplast biogenesis. Nature **403:** 203–207.

Baum, J., A. T. Papenfuss, B. Baum, T. P. Speed and A. F. Cowman, 2006 Regulation of apicomplexan actin-based motility. Nat. Rev. Microbiol. **4:** 621–628.

Becker, T., J. Hritz, M. Vogel, A. Caliebe, B. Bukau *et al.*, 2004a Toc12, a novel subunit of the intermembrane space preprotein translocon of chloroplasts. Mol. Biol. Cell **15:** 5130–5144.

Becker, T., M. Jelic, A. Vojta, A. Radunz, J. Soll *et al.*, 2004b Preprotein recognition by the Toc complex. EMBO J. **23:** 520–530.

Berney, C., and J. Pawlowski, 2006 A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc. Biol. Sci. **273:** 1867–1872.

Bigelow, H. R., D. S. Petrey, J. Liu, D. Przybylski and B. Rost, 2004 Predicting transmembrane beta-barrels in proteomes. Nucleic Acids Res. **32:** 2566–2577.

Butterfield, N. J., 2000 Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology **26:** 386–404.

Caliebe, A., R. Grimm, G. Kaiser, J. Lübeck, J. Soll *et al.*, 1997 The chloroplastic protein import machinery contains a Rieske-type iron-sulfur cluster and a mononuclear iron-binding protein. EMBO J. **16:** 7342–7350.

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

Chen, X., M. Smith, L. Fitzpatrick and D. Schnell, 2002 In vivo analysis of the role of atTic20 in protein import into chloroplasts. Plant Cell **14:** 641–654.

Chew, O., R. Lister, S. Qbadou, J. L. Heazlewood, J. Soll *et al.*, 2004 A plant outer mitochondrial membrane protein with high amino acid sequence identity to a chloroplast protein import receptor. FEBS Lett. **557:** 109–114.

Chigri, F., F. Hörmann, A. Stamp, D. K. Stammers, B. Bölter *et al.*, 2006 Calcium regulation of chloroplast protein translocation is mediated by calmodulin binding to Tic32. Proc. Natl. Acad. Sci. USA **103:** 16051–16056.

Chou, M. L., L. M. Fitzpatrick, S. L. Tu, G. Budziszewski, S. Potter-Lewis *et al.*, 2003 Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. EMBO J. **22:** 2970–2980.

Chou, M. L., C. C. Chu, L. J. Chen, M. Akita and H. M. Li, 2006 Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts. J. Cell Biol. **175:** 893–900.

Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner, 2004 WebLogo: a sequence logo generator. Genome Res. **14:** 1188–1190.

Dávila-Aponte, J., K. Inoue and K. Keegstra, 2003 Two chloroplastic protein translocation components, Tic110 and Toc75, are conserved in different plastid types from multiple plant species. Plant Mol. Biol. **51:** 175–181.

Derelle, E., C. Ferraz, S. Rombauts, P. Rouzé, A. Worden *et al.*, 2006 Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc. Natl. Acad. Sci. USA **103:** 11647–11652.

Duy, D., G. Wanner, A. R. Meda, N. Von Wiren, J. Soll *et al.*, 2007 PIC1, an ancient permease in Arabidopsis chloroplasts, mediates iron transport. Plant Cell **19:** 986–1006.

Eckart, K., L. Eichacker, K. Sohrt, E. Schleiff, L. Heins *et al.*, 2002 A Toc75-like protein import channel is abundant in chloroplasts. EMBO Rep. **3:** 557–562.

Emanuelsson, O., S. Brunak, G. Von Heijne and H. Nielsen, 2007 Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. **2:** 953–971.

Ertel, F., O. Mirus, R. Bredemeier, S. Moslavac, T. Becker *et al.*, 2005 The evolutionarily related beta-barrel polypeptide transporters from Pisum sativum and Nostoc PCC7120 contain two distinct functional domains. J. Biol. Chem. **280:** 28281–28289.

Franzén, L. G., J. D. Rochaix and G. Von Heijne, 1990 Chloroplast transit peptides from the green alga Chlamydomonas reinhardtii share features with both mitochondrial and higher plant chloroplast presequences. FEBS Lett. **260:** 165–168.

Gentle, I. E., L. Burri and T. Lithgow, 2005 Molecular architecture and function of the Omp85 family of proteins. Mol. Microbiol. **58:** 1216–1225.

Gray, J., E. Wardzala, M. Yang, S. Reinbothe, S. Haller *et al.*, 2004 A small family of LLS1-related non-heme oxygenases in plants with an origin amongst oxygenic photosynthesizers. Plant Mol. Biol. **54:** 39–54.

Guindon, S., and O. Gascuel, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:** 696–704.

Heins, L., A. Mehrle, R. Hemmler, R. Wagner, M. Küchler *et al.*, 2002 The preprotein conducting channel at the inner envelope membrane of plastids. EMBO J. **21:** 2616–2625.

Hernández Torres, J., M. A. Maldonado and J. Chomilier, 2007 Tandem duplications of a degenerated GTP-binding domain at the origin of GTPase receptors Toc159 and thylakoidal SRP. Biochem. Biophys. Res. Commun. **364:** 325–331.

Hofmann, N. R., and S. M. Theg, 2003 Physcomitrella patens as a model for the study of chloroplast protein transport: conserved machineries between vascular and non-vascular plants. Plant Mol. Biol. **53:** 621–632.

Hofmann, N. R., and S. M. Theg, 2005 Toc64 is not required for import of proteins into chloroplasts in the moss Physcomitrella patens. Plant J. **43:** 675–687.

Hörmann, F., M. Küchler, D. Sveshnikov, U. Oppermann, Y. Li *et al.*, 2004 Tic32, an essential component in chloroplast biogenesis. J. Biol. Chem. **279:** 34756–34762.

Inaba, T., M. Li, M. Alvarez-Huerta, F. Kessler and D. J. Schnell, 2003 atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. J. Biol. Chem. **278:** 38617–38627.

Inaba, T., M. Alvarez-Huerta, M. Li, J. Bauer, C. Ewers *et al.*, 2005 Arabidopsis tic110 is essential for the assembly and function of the protein import machinery of plastids. Plant Cell **17:** 1482–1496.

Inoue, K., and D. Potter, 2004 The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms. Plant J. **39:** 354–365.

Jackson-Constan, D., and K. Keegstra, 2001 Arabidopsis genes encoding components of the chloroplastic protein import apparatus. Plant Physiol. **125:** 1567–1576.

Jarvis, P., and J. Soll, 2002 Toc, tic, and chloroplast protein import. Biochim. Biophys. Acta **1590:** 177–189.

Käll, L., A. Krogh and E. L. Sonnhammer, 2004 A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. **338:** 1027–1036.

Kallberg, Y., U. Oppermann, H. Jörnvall and B. Persson, 2002 Short-chain dehydrogenases/reductases (SDRs). Eur. J. Biochem. **269:** 4409–4417.

Kouranov, A., and D. J. Schnell, 1997 Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. J. Cell Biol. **139:** 1677–1685.

Kouranov, A., X. Chen, B. Fuks and D. J. Schnell, 1998 Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. J. Cell Biol. **143:** 991–1002.

Kovacheva, S., J. Bédard, R. Patel, P. Dudley, D. Twell *et al.*, 2005 In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. Plant J. **41:** 412–428.

Kozjak, V., N. Wiedemann, D. Milenkovic, C. Lohaus, H. Meyer *et al.*, 2003 An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. J. Biol. Chem. **278:** 48520–48523.

Krogh, A., B. Larsson, G. Von Heijne and E. L. Sonnhammer, 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:** 567–580.

Küchler, M., S. Decker, F. Hörmann, J. Soll and L. Heins, 2002 Protein import into chloroplasts involves redox-regulated proteins. EMBO J. **21:** 6136–6145.

Leister, D., 2003 Chloroplast research in the genomic age. Trends Genet. **19:** 47–56.

Li, M., and D. J. Schnell, 2006 Reconstitution of protein targeting to the inner envelope membrane of chloroplasts. J. Cell Biol. **175:** 249–259.

López-Juez, E., 2007 Plastid biogenesis: between light and shadows. J. Exp. Bot. **58:** 11–26.

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

Martin, W., B. Stoebe, V. Goremykin, S. Hapsmann, M. Hasegawa *et al.*, 1998 Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393:** 162–165.

Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen *et al.*, 2002 Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA **99:** 12246–12251.

McFadden, G. I., and G. G. Van Dooren, 2004 Evolution: red algal genome affirms a common origin of all plastids. Curr. Biol. **14:** R514–R516.

Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz *et al.*, 2007 The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318:** 245–250.

Mokranjac, D., and W. Neupert, 2005 Protein import into mitochondria. Biochem. Soc. Trans. **33:** 1019–1023.

Oppermann, U., C. Filling, M. Hult, N. Shafqat, X. Wu *et al.*, 2003 Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chem. Biol. Interact. **143–144:** 247–253.

Palenik, B., J. Grimwood, A. Aerts, P. Rouzé, A. Salamov *et al.*, 2007 The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. USA **104:** 7705–7710.

Paschen, S., T. Waizenegger, T. Stan, M. Preuss, M. Cyrklaff *et al.*, 2003 Evolutionary conservation of biogenesis of beta-barrel membrane proteins. Nature **426:** 862–866.

Patron, N. J., and R. F. Waller, 2007 Transit peptide diversity and divergence: a global analysis of plastid targeting signals. BioEssays **29:** 1048–1058.

Pujol, C., L. Maréchal-Drouard and A. M. Duchêne, 2007 How can organellar protein N-terminal sequences be dual targeting signals? In silico analysis and mutagenesis approach. J. Mol. Biol. **369:** 356–367.

Qbadou, S., T. Becker, O. Mirus, I. Tews, J. Soll *et al.*, 2006 The molecular chaperone Hsp90 delivers precursor proteins to

the chloroplast import receptor Toc64. EMBO J. **25:** 1836–1847.

Qbadou, S., T. Becker, T. Bionda, K. Reger, M. Ruprecht *et al.*, 2007 Toc64–a preprotein-receptor at the outer membrane with bipartide function. J. Mol. Biol. **367:** 1330–1346.

Rensing, S. A., J. Ick, J. A. Fawcett, D. Lang, A. Zimmer *et al.*, 2007 An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. BMC Evol. Biol. **7:** 130.

Rensing, S. A., D. Lang, A. D. Zimmer, A. Terry, A. Salamov *et al.*, 2008 The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319:** 64–69.

Reumann, S., J. Davila-Aponte and K. Keegstra, 1999 The evolutionary origin of the protein-translocating channel of chloroplastic envelope membranes: identification of a cyanobacterial homolog. Proc. Natl. Acad. Sci. USA **96:** 784–789.

Reumann, S., K. Inoue and K. Keegstra, 2005 Evolution of the general protein import pathway of plastids (review). Mol. Membr. Biol. **22:** 73–86.

Robbens, S., E. Derelle, C. Ferraz, J. Wuyts, H. Moreau *et al.*, 2007 The complete chloroplast and mitochondrial DNA sequence of Ostreococcus tauri: organelle genomes of the smallest eukaryote are examples of compaction. Mol. Biol. Evol. **24:** 956–968.

Sánchez-Pulido, L., D. Devos, S. Genevrois, M. Vicente and A. Valencia, 2003 POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. Trends Biochem. Sci. **28:** 523–526.

Schleiff, E., M. Jelic and J. Soll, 2003 A GTP-driven motor moves proteins across the outer envelope of chloroplasts. Proc. Natl. Acad. Sci. USA **100:** 4604–4609.

Schnell, D., F. Kessler and G. Blobel, 1994 Isolation of components of the chloroplast protein import machinery. Science **266:** 1007–1012.

Sohrt, K., and J. Soll, 2000 Toc64, a new component of the protein translocon of chloroplasts. J. Cell Biol. **148:** 1213–1221.

Soll, J., and E. Schleiff, 2004 Protein import into chloroplasts. Nat. Rev. Mol. Cell Biol. **5:** 198.

Sommer, M. S., S. B. Gould, P. Lehmann, A. Gruber, J. M. Przyborski *et al.*, 2007 Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol. Biol. Evol. **24:** 918–928.

Stahl, T., C. Glockmann, J. Soll and L. Heins, 1999 Tic40, a new "old" subunit of the chloroplast protein import translocon. J. Biol. Chem. **274:** 37467–37472.

Sveshnikova, N., J. Soll and E. Schleiff, 2000 Toc34 is a preprotein receptor regulated by GTP and phosphorylation. Proc. Natl. Acad. Sci. USA **97:** 4973–4978.

Teng, Y., Y. Su, L. Chen, Y. Lee, I. Hwang *et al.*, 2006 Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. Plant Cell **18:** 2247–2257.

Timmis, J. N., M. A. Ayliffe, C. Y. Huang and W. Martin, 2004 Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. **5:** 123–135.

Tranel, P., and K. Keegstra, 1996 A novel, bipartite transit peptide targets OEP75 to the outer membrane of the chloroplastic envelope. Plant Cell **8:** 2093–2104.

Tripp, J., K. Inoue, K. Keegstra and J. E. Froehlich, 2007 A novel serine/proline-rich domain in combination with a transmembrane domain is required for the insertion of AtTic40 into the inner envelope membrane of chloroplasts. Plant J. **52:** 824–838.

Voigt, A., M. Jakob, R. Klösgen and M. Gutensohn, 2005 At least two Toc34 protein import receptors with different specificities are also present in spinach chloroplasts. FEBS Lett. **579:** 1343–1349.

Vojta, A., M. Alavi, T. Becker, F. Hörmann, M. Küchler *et al.*, 2004 The protein translocon of the plastid envelopes. J. Biol. Chem. **279:** 21401–21405.

Waegemann, K., and J. Soll, 1991 Characterization of the protein import apparatus in isolated outer envelopes of chloroplasts. Plant J. **1:** 149–158.

Weibel, P., A. Hiltbrunner, L. Brand and F. Kessler, 2003 Dimerization of Toc-GTPases at the chloroplast protein import machinery. J. Biol. Chem. **278:** 37321–37329.

Yeh, Y., M. Kesavulu, H. Li, S. Wu, Y. Sun *et al.*, 2007 Dimerization is important for the GTPase activity of chloroplast translocon components atToc33 and psToc159. J. Biol. Chem. **282:** 13845–13853.

Zhang, X. P., and E. Glaser, 2002 Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. Trends Plant Sci. **7:** 14–21.

Communicating editor: O. Vallon