the FEBS Journal

The chloroplast ClpP complex in *Chlamydomonas reinhardtii* contains an unusual high molecular mass subunit with a large apical domain

Wojciech Majeran[1], Giulia Friso[2], Klaas Jan van Wijk[2] and Olivier Vallon[1]

1 Institut de Biologie Physico-Chimique, Paris, France
2 Department of Plant Biology Cornell University, Ithaca, New York, USA

Keywords
mass spectroscopy; native gel electrophoresis; protein complex; proteolysis; Volvocale

Correspondence
O. Vallon, UMR 7141, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France
Fax: +33 15841 5022
Tel: +33 15841 5058
E-mail: ovallon@ibpc.fr

(Received 15 July 2005, revised 23 August 2005, accepted 31 August 2005)

doi:10.1111/j.1742-4658.2005.04951.x

The composition of the chloroplast-localized protease complex, ClpP, from the green alga *Chlamydomonas reinhardtii* was characterized by nondenaturing electrophoresis, immunoblotting and MS. The detected ClpP complex has a native mass of ≈ 540 kDa, which is ≈ 200 kDa higher than ClpP complexes in higher plant chloroplasts, mitochondria or bacteria. The 540-kDa ClpP complex contains two nuclear-encoded ClpP proteins (ClpP3 and P5) and five ClpR (R1, R2, R3, R4 and R6) proteins, as well two proteins, ClpP1_L and ClpP1_H, both probably derived from the plastid *clpP1* gene. ClpP1_H is 59 kDa and contains a ≈ 30-kDa insertion sequence (IS1) not found in other ClpP proteins, responsible for the high MW of the complex. Based on comparison with other sequences, IS1 protrudes as an additional domain on the apical surface of the ClpP/R complex, probably preventing interaction with the HSP100 chaperone. ClpP1_L is a 25-kDa protein similar in size to other ClpP proteins and could arise by post-translational processing of ClpP1_H. Chloramphenicol-chase experiments show that ClpP1_L and ClpP1_H have a similar half-life, indicating that both are stable components of the complex. The structure of the ClpP complex is further discussed in conjunction with a phylogenetic analysis of the ClpP/R genes. A model is proposed for the evolution of the algal and plant complex from its cyanobacterial ancestor.

Most intracellular proteolysis is carried out by large self-compartmentalized ATP-dependent proteases, combining a chaperone and a peptidase activity [1]. The chaperone activity is necessary to unfold protein substrates, and feed them into a proteolytic chamber where peptidolysis occurs. In the eukaryotic cell, mitochondria and chloroplasts harbor three major types of ATP-dependent proteases, all inherited from their eubacterial ancestors: FtsH, Lon, and Clp [2–4].

Clp proteases are composed of two components, a chaperone of the Hsp100 type and a peptidase of the ClpP family. In *Escherichia coli* and other bacteria, the chaperone is either ClpA or ClpX, resulting in the formation of ClpAP, ClpXP or mixed ClpAXP complexes [3,5–7]. ClpP is a serine-type endopeptidase, with the three catalytic residues, Ser, His and Asp appearing in this order in the sequence [8]. The X-ray crystallographic structure of the *E. coli* ClpP complex [9] shows two stacked rings of seven identical ClpP subunits, delineating an internal proteolytic chamber where the 14 catalytic sites are exposed. By itself, this tetradecameric complex is only capable of hydrolyzing short peptides. Degradation of proteins requires association with ClpA or ClpX (which also govern substrate specificity) and ATP hydrolysis [10,11]. The chaperones can function on their own to remodel or refold denatured proteins, but their association with ClpP promotes a different mechanism, whereby the extended polypeptide chain of the substrate is fed into the proteolytic ClpP chamber via a narrow axial opening. In *E. coli*, the ClpAP and ClpXP protease complexes function in the degradation of denatured proteins, especially under heat stress, but also of specific substrates recognized via N- or C-terminal sequence motifs. The diversity of these motifs was revealed when a tagged and inactive variant of *E. coli* ClpP was used to trap substrates of ClpXP *in vivo* [12]. In addition, ClpXP participates in the degradation of nascent proteins stalled on ribosomes, after they are C-terminally tagged and released by the *ssrA trans*-termination system. Substrate binding is influenced by helper and modulator proteins, such as SspB [13] or ClpS [14].

Much less is known about the Clp proteases of eukaryotes. The mitochondrial ClpP complex appears highly similar to the bacterial enzyme [15]. It interacts with a ClpX chaperone to form an ATP-dependent protease [15–17]. Plant mitochondrial ClpP2 has been shown to form a homo-oligomeric complex of ≈ 320 kDa [18]. In plastids, the most likely partners of ClpP are the ClpC chaperone [19] and ClpD (Erd1) [20]. The ClpP complex in plastids of *Arabidopsis thaliana* and other Brassicaceae has been examined in detail. It is a hetero-oligomer, slightly larger than its mitochondrial counterpart ($\approx$350 kDa), associating nucleus- and plastid-encoded subunits. In all plastid types examined, it associates five different ClpP proteins (ClpP1, ClpP3, ClpP4, ClpP5 and ClpP6, of which ClpP1 is chloroplast-encoded), and four nonproteolytic ClpR proteins (ClpR1, ClpR2, ClpR3 and ClpR4) [18,21]. ClpR proteins are homologous to ClpP, but lack one or several of the catalytic site residues, and are therefore supposed to play a structural, rather than a catalytic, role in the complex. In addition, the plastid ClpP/R complex contains two additional subunits unique to land plants, ClpS1 and ClpS2. The functions of ClpS1,2 (which are not homologous to the *E. coli* ClpS) are unknown, but molecular modeling based on their homology to the N-terminal domain of ClpA suggests that they dock onto the apical surface of the (presumably tetradecameric) ClpP/R complex to regulate association with the chaperone [18].

All of the plant Clp proteins are encoded in the nucleus, except for ClpP1 which is plastid-encoded in green algae and vascular plants, i.e. the green lineage of plants. ClpP genes are also found in Cyanobacteria [22] and in the genome of the Cyanophora cyanelle, an ancestral chloroplast. In the green alga *Chlamydomonas reinhardtii* and in vascular plants, the plastid *clpP1* gene is essential and cannot be disrupted [23,24]. Reducing its expression level by mutating its initiation codon leads to a reduction in degradation rate for several thylakoid membrane proteins under stress conditions or in the presence of destabilizing mutations [25,26]. Other than that, little is known about the substrates and functions of the plastid Clp protease.

ClpP1 proteins in the *Chlamydomonas* genus are unusual in that they contain insertion sequences, not found in other ClpP proteins, which have been proposed to behave as protein introns [23]. One of them, IS2, is found only in the species *C. eugametos*, and possesses characteristics of the well-known self-splicing protein elements called inteins [27–29]. Inteins are autonomously folding protein domains capable of self-excision, and are present in many essential proteins throughout the three kingdoms of life. In contrast, the other insertion sequence, IS1, which is found in both *C. eugametos* and *C. reinhardtii*, lacks most of the typical features of an intein, in particular the LAGLI/DADG motif and C-terminal HN dipeptide [23]. In a previous study, we have shown that antibodies raised to the entire *C. reinhardtii* ClpP1 reading frame recognize two proteins of 25 kDa (referred to here as $\mathrm{ClpP1_L}$) and 59 kDa ($\mathrm{ClpP1_H}$). As splicing of the *clpP1* mRNA has been ruled out [23]; unpublished data), this supports the protein splicing hypothesis. Presumably, $\mathrm{ClpP1_H}$ is the primary translation product of *clpP1*, while $\mathrm{ClpP1_L}$ is derived from $\mathrm{ClpP1_H}$ by splicing or some other form of post-translational processing. Here, we analyze the ClpP/R complex of *C. reinhardtii* by gel electrophoresis followed by MS and show that it contains both $\mathrm{ClpP1_L}$ and $\mathrm{ClpP1_H}$.

Results

Identification of nuclear ClpP/R genes

The *C. reinhardtii* EST databases contained sequences from eight nuclear-encoded ClpP/R homologs (Table 1). By combining EST data and partial sequencing of selected cDNAs, complete cDNA sequences were obtained for all of them except for *CLPR3* and *CLPP2*. All the corresponding *CLPP/R* genes were found in the *C. reinhardtii* draft genomic sequence, version 2.0, some still containing sequence gaps. Models for four of the genes were corrected based on EST data, in-house cDNA sequencing, comparison with version 1.0 of the genome, or alignment with plant orthologs (see Table 1; the proposed changes have been deposited as model notes in the JGI gene models). The *C. reinhardtii* ClpP/R proteins and genes were named on the basis of their closest *Arabidopsis* homologs [2], as judged from the phylogenetic tree (Fig. 1) deduced from a CLUSTALW alignment of ClpP proteases (Fig. S1). Only the central domain for each gene was used for tree building, avoiding N-terminal and C-terminal extensions where alignment was not meaningful.

Only three of the *C. reinhardtii* proteins (ClpP2, ClpP4 and ClpP5) are predicted to contain the three conserved catalytic site residues, and thus to be enzymatically active. The other five were missing either

Table 1. ClpP/R genes in *Chlamydomonas reinhardtii.* For each gene are indicated the name, best contig in EST assembly (Chlre2), gene model(s) in version 2.0 of the *Chlamydomonas* genome, predicted molecular mass and length of precursor and mature protein, N-terminal sequence, length of C-terminal extension and peptides found in the ESI/MS/MS analysis.

Protein	Best contig in EST assembly[a]	Additional sequencing	Gene model	Precursor (kDa); length (AA)	Mature: (kDa); length (AA)	Precursor N-terminal sequence	Length of C-terminal extension[b]	MS sequence tags	Notes on protein
ClpP1	–	–	CLPP_CHLRE (chloroplast)	59.3 [523]	26.7 [237]?	MPIGVPRII. . .	41 (–2)	LDVAEIYSLSTYRASLVGDSTQTQESNS	Probable post-translational processing or splicing
ClpP2	16.56.2.51	–	C_160152[c]	25,0 [231]	21,3 [194]	MQRLSALAA. . .	7 (18)	None	
ClpP4	37.11.2.11	AV387965	C_370132[d]	38.4 [345]	32.8 [293]	MVAAALLGG. . .	105 (38–49)	TAEFQGDPMGLLLRYGLIDHIIGGEEAVFNVKRNSMPNSR	Long C-terminal extension
ClpP5	162.16.2.11	AV387186	C_1620010	27.8 [256]	22.0 [201][e]	MAQLLLQNK. . .	11 (9)	FQGVVSQLFQQRGPPPPNLPVIER	
ClpR1	24.46.1.0	AV619744	C_240023[f]	46.2 [411]	38.9 [345]?	MLLNRPKLG. . .	31 (31)	FGNPPDLPSLLLQQRAPLYTGVTWK AVDAQLQANELDYATKPLFLPEAER	Long N-terminal extension
ClpR2	56.23.1.51	–	C_560017	31.6 [282]	25.0 [220]	MQALNQRPS. . .	22 (15)	None	
ClpR3	105.11.2.11	–	(C_1050047, C_17380001, C_24440001)[g]	46.5 [415]	43.4 [386]	MRVHHAMTG. . .	178 (23)	SLPHSLAMIQQPRDGVKLAILNAE. . .	Long C-terminal extension
ClpR4	12.4.2.11	–	C_120194	32.7 [293]	28.9 [257]	MAALGCLSR. . .	47 (15)	LGGMQASDIDIYRFGNEHEAIAVYSMMKEAGPPPDLATR TEEQIMTDFTRPRHEAIAVYSMMK	
ClpR6	**19.134.4.11**	AV622513	C_190082 [h]	31.2 [283]	25.6 [231]	MATLLQHGR. . .	32 (0)	EVGLVDDLTPGPFLKIIYINDKLEGLIDEIIR LMMTQPMGGSQGDIYQIK	

[a] In bold if believed to be correct. [b] Compared to *E. coli* ClpP; value for the *Arabidopsis* orthologue(s) is indicated in parentheses. [c] An additional 5′ exon and intron must be introduced to generate the full-length protein; the unique, truncated EST from this gene is from an incompletely matured mRNA. [d] Sequence gap; exon 5 wrong. [e] Based on MS/MS data; 21 g and C9 nucleotide sequences differ but produce the same protein. [f] Sequence gap; exon 4 wrong. [g] Sequence gaps and misassembly; corrected based on cDNA data and genome version 1.0 : 5′ must be extended, and two consecutive frameshifts corrected, restoring conserved residues in helix 5; 10th and 11th exons wrong. [h] The splice sites of the eighth intron must be shifted seven nucleotides downstream, restoring two conserved residues.

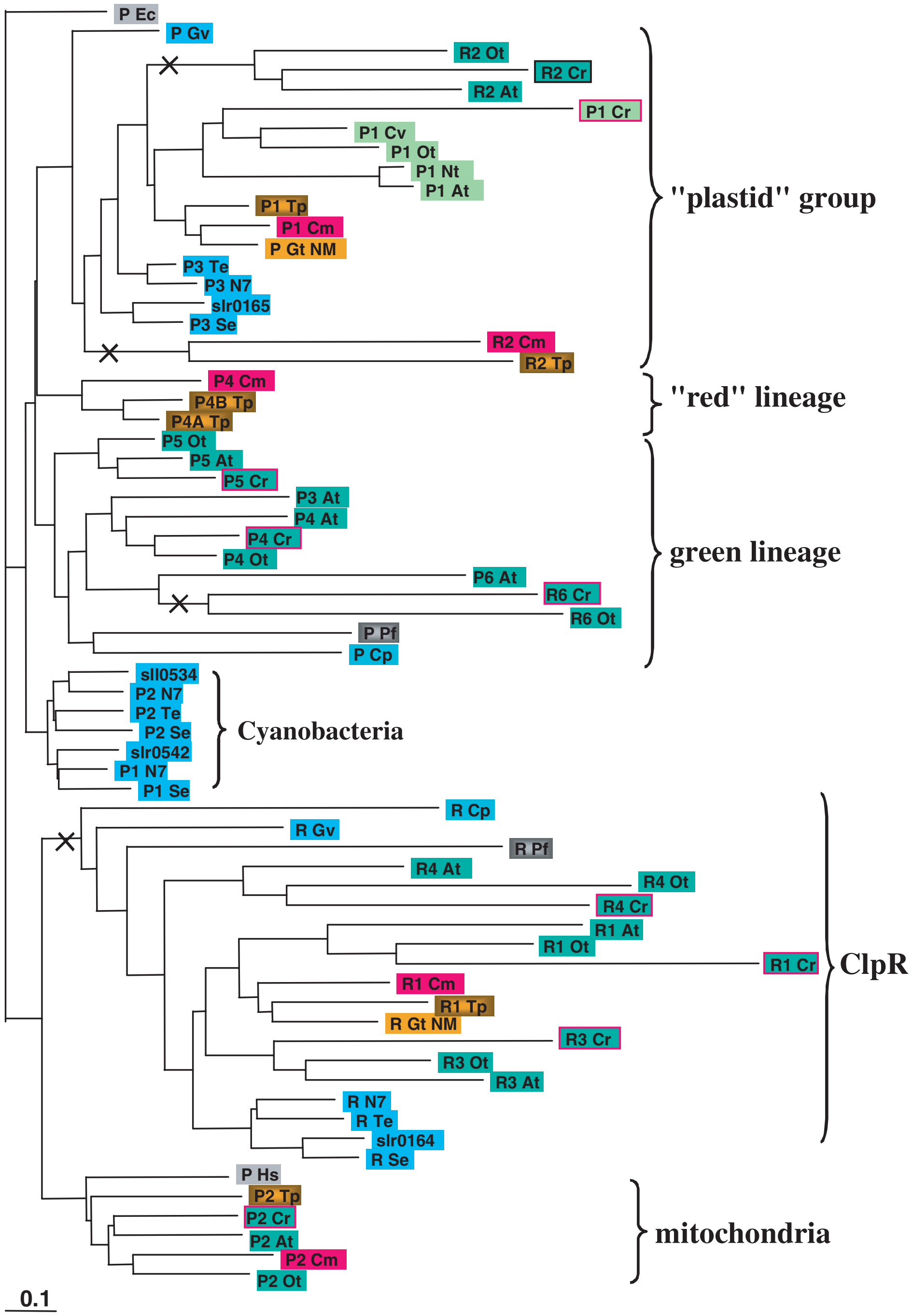


Fig. 1. Phylogenetic tree, obtained by the Neighbor-Joining method (with Kimura's distance correction) of ClpP/R proteins from *Chlamydomonas* (boxed in red) and other photosynthetic eukaryotes and prokaryotes. Blue color indicates Cyanobacteria: *Gleobacter violaceus* (Gv), *Nostoc* sp. PCC 7120 (N7), *Synechocystis* sp. PCC 6803 (S6), *Synecococcus elongatus* (Se) and *Thermosynechococcus elongatus* (Te). Greenish blue indicates *Cyanophora paradoxa* (Cp, a Glaucocystophyte) and red *Cyanidioschyzon merolae* (Cm, a Rhodophyte). The brown color indicates *Guillardia theta* (Gt, a Cryptophyte) and *Thalassiosira pseudonana* (Tp, a Diatom). The Viridiplantae (green, of a lighter shade for chloroplast genes) are *Arabidopsis thaliana* (At), *Chlamydomonas reinhardtii* (Cr), *Chlorella vulgaris* (Cv), *Nicotiana tabaccum* (Nt), *Ostreococcus tauri* (Ot). Non photosynthetic organisms (gray) are *E. coli* (Ec), *Plasmodium falciparum* (Pf) and Homo sapiens (Hs). Genes that have undergone two migration events due to secondary endosymbiosis are shown with shading. The branches leading to ClpR proteins, where loss of catalytic activity is presumed to have occurred, are marked by a cross. The alignment used (Fig. S1) was excerpted from a large-scale alignment of ClpP from plants and selected bacteria (available at EMBL as ALIGN_000912), after truncation to the ClpP domain. *E. coli* ClpP was used as the outgroup.

one, two, or three of the catalytic residues S, H and D and were named ClpR1, ClpR2, ClpR3, ClpR4 and ClpR6, respectively. The latter is closest to the *A. thaliana* ClpP6 protein, but the catalytic H is missing in *Chlamydomonas* ClpR6, as well as in orthologs found in two other green algae, *Volvox carteri* and *Ostreococcus tauri*. Compared to *E. coli* ClpP, several of the *C. reinhardtii* proteins showed long C-terminal extensions, usually longer than the *A. thaliana* homologs (Table 1): C-terminal extensions have been proposed to fold back onto the apical surface of the complex, potentially blocking the chaperone binding site [15,18]. For all of the nuclear-encoded ClpP/R proteins, the presence of an N-terminal extension when compared to bacterial homologs suggests that they are targeted to an organelle. The TARGETP and PREDOTAR programs which are used to predict intracellular sorting in vascular plants confirmed targeting to plastids or mitochondria; however, it is difficult to predict which, as the programs do not discriminate well between *Chlamydomonas* plastid and mitochondrial targeting sequences. In addition, the *Chlamydomonas* ClpR1 contains a particularly long N-terminal extension (168 residues before the part conserved with other ClpP), predicted to be retained in the mature protein. A similar extension is also found in *Arabidopsis* but no significant similarity can be identified between the *Chlamydomonas* and *Arabidopsis* ClpR1 proteins in this region.

Importantly, none of the nuclear-encoded ClpP/R proteins of *Chlamydomonas* showed extended similarity to the chloroplast-encoded ClpP1. No stretch of similarity longer than five residues was observed, making it unlikely that the ClpP1_L band recognized by the antibodies corresponds to the product of one of these nuclear genes.

Experimental identification of the *Chlamydomonas* ClpP/R complex

As a first step towards the characterization of the *C. reinhardtii* ClpP/R complex, we analyzed the subcellular localization of *clpP1* gene products. Purified chloroplasts from the cell wall-less mutant CW15 were fractionated into a stroma and a crude membrane fraction (thylakoids plus envelope). As expected, Western blotting identified ClpP1_L mostly in the stromal fraction, with a small proportion (5–10%) associated with the membranes (Fig. 2). A similar proportion was found in membranes purified by floatation on a sucrose gradient (not shown). When chloroplast membranes were further treated with the Yeda press (Fig. 2, lane 4), the bound ClpP1 was efficiently released, indicating a loose association similar to that observed for higher plant ClpP [18,21]. A similar behavior was found for ClpP1_H and ClpC (data not shown), as well as for the chloroplast GrpE homolog CGE1 and for RbcL, both of which can be taken as examples of stromal proteins.

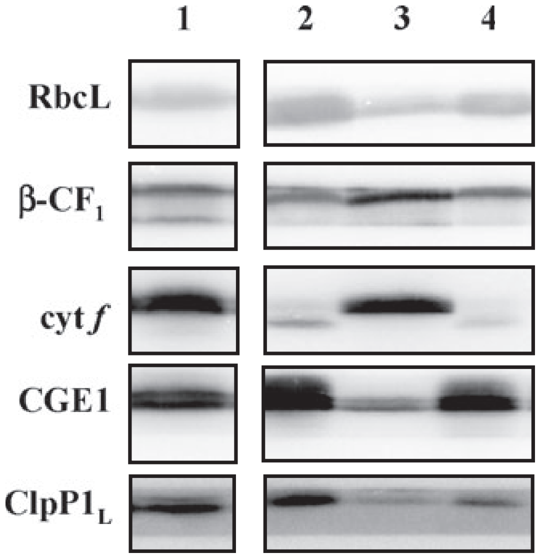


Fig. 2. Chloroplast localization of ClpP1. Immunoblots were reacted with antibodies to ClpP1, CGE1 (chloroplast GrpE homolog), cytochrome *f*, the β subunit of the ATP-synthase CF$_1$, and RbcL, the large subunit of Rubisco. 1: Chloroplasts isolated from strain CW15; 2: soluble fraction and 3: membrane fraction after mechanical lysis of chloroplasts; 4: soluble fraction after Yeda press treatment of fraction 3 (overloaded approximately 10-fold).

Total soluble proteins, prepared by French press lysis of wild-type cells were separated by nondenaturing electrophoresis (colorless-native; CN/PAGE) on a 4–13% gradient gel and blotted onto nitrocellulose for immunodetection of ClpP1. In these experiments, a high molecular mass ClpP complex could be identified, migrating just below the 550-kDa Rubisco complex (Fig. 3). Based on comparison with molecular mass standards and the chloroplast proteins CF$_1$ (390 kDa) and Rubisco (550 kDa), the apparent molecular mass of the ClpP complex is ≈ 540 kDa. This is markedly higher than the 350-kDa complex identified in higher plants. No ClpP1 was detected in the low molecular mass region of the gel. After electrophoresis in the second dimension, many proteins could be resolved but, due the complexity of the protein mixture, no particular silver-stained spot could be identified as candidate components of the ClpP1-immunoreactive complex. When we used, instead of a whole cell lysate, chloroplast stromal proteins or proteins released from the membranes by Yeda press treatment, we encountered the same caveat. The only MS/MS peptides that could

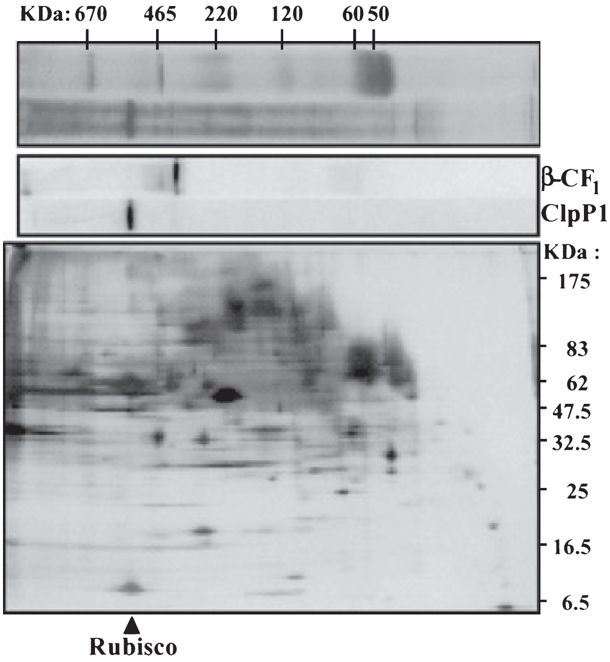


Fig. 3. Identification of the ClpP/R complex. For clarity, native gels are shown horizontally, migration left to right, and denaturing gels (SDS or SDS-urea) are shown vertically, migration top to bottom. Top panel: nondenaturing 4–13% gradient CN/PAGE (Coomassie blue staining), showing molecular mass markers (bovine thyroglobulin, 670 kDa; *E. coli* β-galactosidase, 465 kDa; bovine catalase, 220 kDa; BSA, 120 and 60 kDa; apyrase, 50 kDa) and soluble proteins from wild-type cells ruptured in a French press. Middle panel: immunoblot of the CN/PAGE with antibodies to β-CF$_1$ and ClpP1. Lower panel: second dimension SDS/PAGE (7.5–15% acrylamide gradient, silver staining; the size of the molecular mass markers is indicated on the right, the position of Rubisco is shown by an arrowhead).

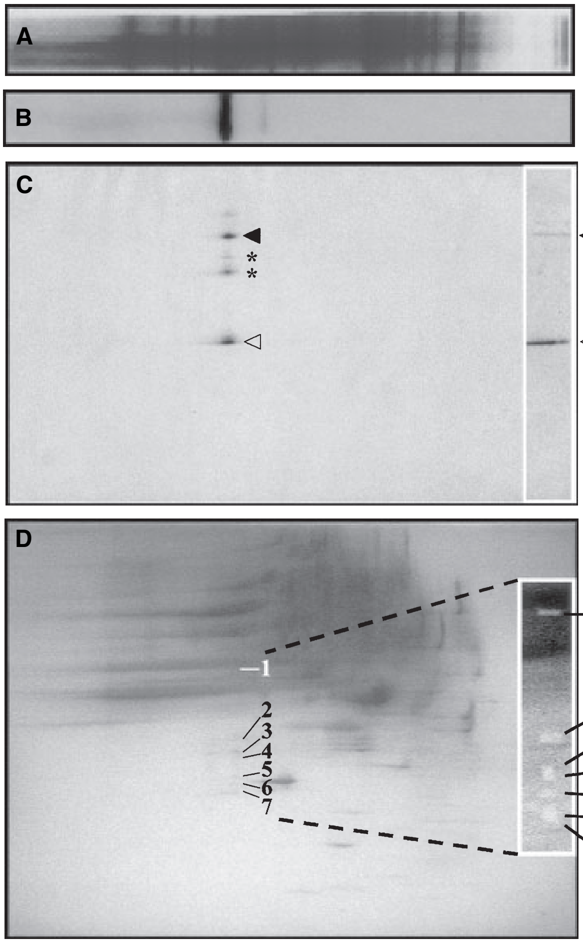


Fig. 4. Identification of the subunits of the ClpP/R complex. Chloroplasts from the *cw15 rbcL18–5b* strain were ruptured by Yeda press treatment, and soluble proteins separated by 2D-electrophoresis. (A) Coomassie blue staining, and (B) immunoblot with ClpP1 antibody, of the first dimension (CN/PAGE, 4–18% acrylamide gradient). (C) Immunoblot with ClpP1 antibody of the second dimension (SDS/urea PAGE, 12–18% acrylamide gradient); on the right is an immunoblot of the same fraction separated by SDS/urea PAGE, with arrowheads showing the positions of ClpP1$_H$ and ClpP1$_L$ (filled and open, respectively; * indicates putative degradation products of ClpP1$_H$). (D) Coomassie blue staining of the 2D gel used to cut out the spots for MS/MS. They are labeled 1–7 and an enlargement of the ClpP region after bands were cut out is shown on the right.

be identified in these experiments were from degradation products of the very abundant Rubisco large subunit migrating approximately at the same position in the first dimension.

We therefore resorted to using a Rubisco-less mutant as our starting material for chloroplast preparation. Figure 4 shows 2D electrophoresis of a soluble fraction prepared by Yeda press treatment of chloroplasts purified from a double mutant *rbcL-18-5b cw15*. In the first dimension (CN/PAGE, Fig. 4A, B), the ClpP1 antibody again recognizes the high molecular mass complex of ≈ 540 kDa described above. Its subunits are separated according to their size in the second, denaturing, dimension. As can be seen by comparison of Fig. 4B and C, both the ClpP1$_H$ and ClpP1$_L$ forms of ClpP1 are present in the complex, indicating that both are *bona fide* constituents of the ClpP complex. In addition to these major spots, two weaker immunoreactive spots (labeled *) can be seen on Fig. 4C. These spots have been detected occasionally with antibodies raised against different preparations of ClpP1, both in 2D gels and, less frequently, in 1D denaturing gels. We suspect that they represent degradation products of ClpP1$_H$.

In a Coomassie blue stained gel, several protein spots (Fig. 4D, labeled 1–7) appeared to comigrate with the ClpP complex detected by immunoblot. The lower spots migrate approximately in the position of ClpP1$_L$ in this type of gel, just below the LHCII protein P17.

The seven spots were excised from the gel and subjected to in-gel trypsin digestion. The resulting peptides were eluted and analyzed by MALDI-TOF and ESI-Q-TOF MS/MS. All of these spots were found to contain ClpP/R proteins (Table 1), as well as unrelated proteins. Each ClpP/R protein was identified by two to five peptides The only ambiguous peptide was one of those attributed to ClpR3 (DGVKLAILNAE…): although the sequence determined matches ClpR3, the remaining C-terminal peptide mass, 195.14 Da, was not compatible with that predicted from the cDNA or genomic data (.CYER). It may represent a post-translationally processed form of the protein. Spots 1 and 2 contained peptides derived from ClpR3 and ClpR4, respectively. Spots 3–7 contained peptides from ClpR1, ClpR6, ClpP4 and ClpP5, respectively. For ClpP5, one of the sequence tags started after a Q55, not after an R or K. It probably corresponds to the original N-terminal end of the mature protein. Importantly, two peptides derived from ClpP1 were found in spots 4 and 6, whose position corresponds approximately to that of the immunoreactive band ClpP1$_L$ in denaturing gel. Both of them originated from the last 90 C-terminal residues of the protein, one of them from the very C terminus.

In the MS analysis, no peptides were found corresponding to ClpP2. This was expected, based on the well-established mitochondrial localization of the ClpP2 homolog in higher plants. More surprising was the absence of detectable ClpR2, since its *Arabidopsis* homolog has been found consistently in the chloroplast ClpP complex [18,21]. Using antibodies raised against a synthetic peptide from the C terminus of *Chlamydomonas* ClpR2, we found that this protein was present in reduced amounts in the *clpP1-AUU* mutant (Fig. 5A), where the accumulation of the complex is reduced due to a mutation in the start codon [25]. In CN/PAGE, immunoreactivity comigrated exactly with the chloroplast ClpP complex detected with the ClpP1 antibody (Fig. 5B).

IS1 is unique to ClpP1 in *Chlamydomonas* spp. and related organisms

ClpP1$_H$ is the highest molecular mass subunit ever found in a ClpP complex, and one of the most peculiar as its large size is due to the presence of an intervening sequence IS1. IS1 has been reported in several *Chlamydomonas* species, and its sequence has been published for both *C. reinhardtii* and *C. eugametos* [23]. We asked whether other organisms than *Chlamydomonas* contain sequences similar to IS1 in one of their ClpP genes. BLAST searches in the NR and other databases with the two available IS1 sequences failed to detect any other homolog, confirming its very narrow taxonomic range. In particular, *clpP1* from other green algae (*Nephroselmis olivacea*, *Ostreococcus tauri*) were found to contain no IS1-like sequence. To determine the sequence of the *clpP1* gene from the related alga *V. carteri*, a 4.2 kbp sequence contig was assembled from 52 whole genome sequencing reads (available from the JGI website). It encodes a protein of 530 residues with an IS1 sequence in frame with the rest of the protein. The N- and C-terminal regions are virtually identical to those of *Chlamydomonas*. The *Volvox* IS1 is less conserved, but aligns well with its *Chlamydomonas* counterparts (Fig. 6). The best-conserved regions are the N-terminal and C-terminal borders, rich in K and E residues, and two internal regions rich in aromatic residues. The secondary structure prediction algorithms PREDATOR [30] and GOR4 [31] propose a mostly α-helical structure, in particular in the regions that are best conserved (Fig. 6). Efforts to compute a structural model for IS1 (in collaboration with D. Ripoll, Cornell University) have failed

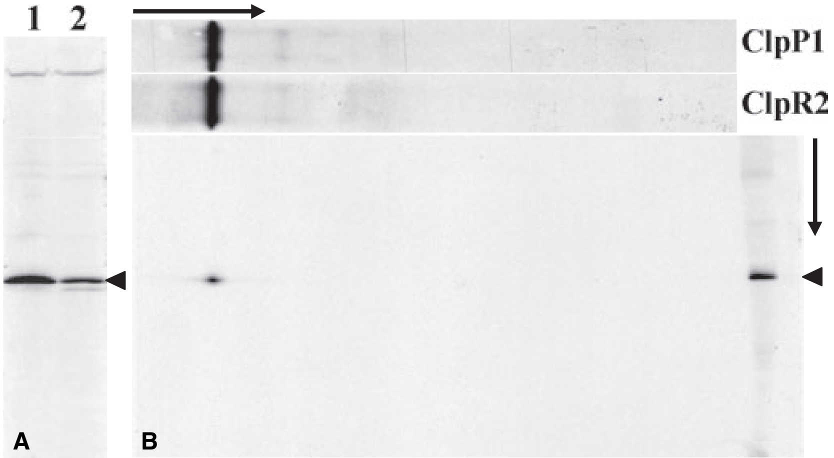
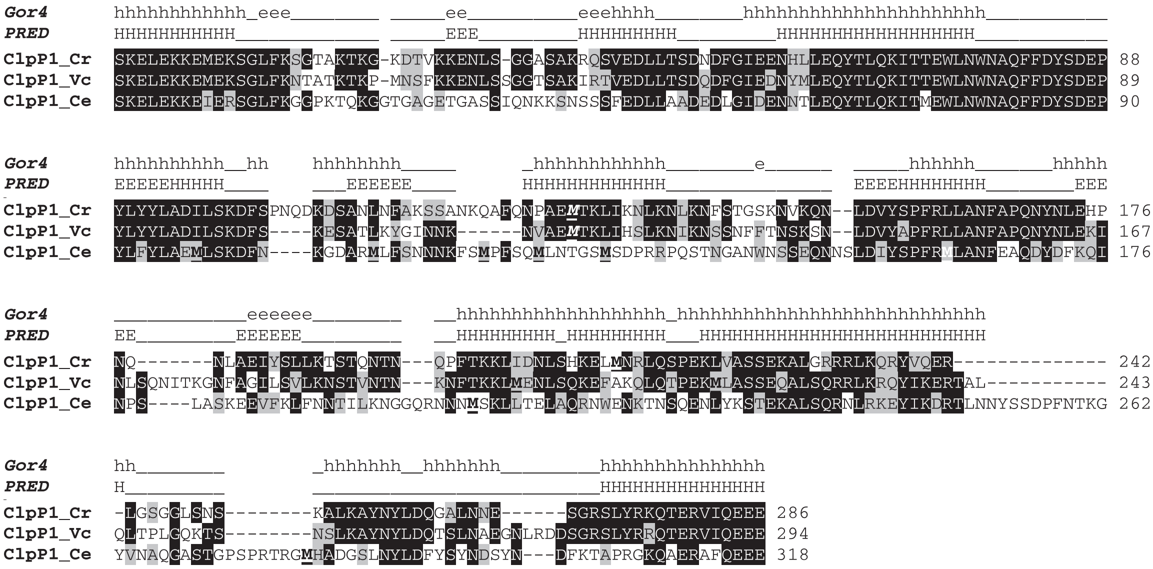


Fig. 5. ClpR2 is part of the ClpP complex. (A) immunoblots of wild-type and *clpP1-AUU* mutant cells, reacted with the ClpR2 antibody. (B) 1D and 2D immunoblots (CN/PAGE, 6–13% acrylamide gradient and SDS/urea/PAGE) of the chloroplast stroma of wild-type cells; on the right, for alignment, wild-type cells have been deposited for electrophoresis in the second dimension.



```
Gor4      hhhhhhhhhhhh_eee________ _____ee__________eeehhhh________hhhhhhhhhhhhhhhhhhhhhh___________
PRED      HHHHHHHHHHH_____________ _____EEE_________HHHHHHHHH_________HHHHHHHHHHHHHHHHHH____________
ClpP1_Cr  SKELEKKEMEKSGLFKSGTAKTKG-KDTVKKENLS-GGASAKRQSVEDLLTSDNDFGIEENHLLEQYTLQKITTEWLNWNAQFFDYSDEP 88
ClpP1_Vc  SKELEKKEMEKSGLFKNTATKTKP-MNSFKKENLSSGGTSAKIRTVEDLLTSDQDFGIEDNYMLEQYTLQKITTEWLNWNAQFFDYSDEP 89
ClpP1_Ce  SKELEKKEIERSGLFKGGPKTQKGGTGAGETGASSIQNKKSNSSSFEDLLAADEDLGIDENNTLEQYTLQKITMEWLNWNAQFFDYSDEP 90

Gor4      hhhhhhhhhh__hh    hhhhhhhh_____      _hhhhhhhhhhhh________e______  _____hhhhhh_______hhhhh
PRED      EEEEEHHHHH____    ___EEEEEE____      HHHHHHHHHHHHH_______________  EEEEHHHHHHHH________EEE
ClpP1_Cr  YLYYLADILSKDFSPNQDKDSANLNFAKSSANKQAFQNPAEMTKLIKNLKNLKNFSTGSKNVKQN--LDVYSPFRLLANFAPQNYNLEHP 176
ClpP1_Vc  YLYYLADILSKDFS----KESATLKYGINNK------NVAEMTKLIHSLKNIKNSSNFFTNSKSN--LDVYAPFRLLANFAPQNYNLEKI 167
ClpP1_Ce  YLFYLAEMLSKDFN----KGDARMLFSNNNKFSMPFSQMLNTGSMSDPRRPQSTNGANWNSSEQNNSLDIYSPFRMLANFEAQDYDFKQI 176

Gor4      ____________eeeeee________   __hhhhhhhhhhhhhhhhhhh_hhhhhhhhhhhhhhhhhhhhhhhhhhhh
PRED      EE_________EEEEEE_________   __HHHHHHHHH_HHHHHHHHH___HHHHHHHHHHHHHHHHHHHHHHHHHH
ClpP1_Cr  NQ-------NLAEIYSLLKTSTQNTN---QPFTKKLIDNLSHKELMNRLQSPEKLVASSEKALGRRRLKQRYVQER-------------- 242
ClpP1_Vc  NLSQNITKGNFAGILSVLKNSTVNTN---KNFTKKLMENLSQKEFAKQLQTPEKMLASSEQALSQRRLKRQYIKERTAL----------- 243
ClpP1_Ce  NPS----LASKEEVFKLFNNTILKNGGQRNNNMSKLLTELAQRNWENKTNSQENLYKSTEKALSQRNLRKEYIKDRTLNNYSSDPFNTKG 262

Gor4      hh________        _hhhhhhh__hhhhhhh_________hhhhhhhhhhhhhhh
PRED      H_________        __________________________HHHHHHHHHHHHHHH
ClpP1_Cr  -LGSGGLSNS--------KALKAYNYLDQGALNNE------SGRSLYRKQTERVIQEEE 286
ClpP1_Vc  QLTPLGQKTS--------NSLKAYNYLDQTSLNAEGNLRDDSGRSLYRRQTERVIQEEE 294
ClpP1_Ce  YVNAQGASTGPSPRTRGMHADGSLNYLDFYSYNDSYN---DFKTAPRGKQAERAFQEEE 318
```

Fig. 6. Alignment of IS1 sequences in *C. reinhardtii*, *V. carteri* and *C. eugametos*. Conserved residues are shaded. The top lines are the secondary structure predictions of the programs GOR4 and PSI-PRED for the *Volvox* IS1.

to generate a reliable prediction, due to the lack of homology with proteins of known structure.

Stability of ClpP1$_H$

If ClpP1$_H$ is the precursor of ClpP1$_L$, then its presence in the ClpP/R complex implies either that the conversion to ClpP1$_L$ is a slow process, or that it is rapid but limited to a fraction of the ClpP1$_H$ produced. To address this question, we examined the stability of the two immunoreactive forms after addition of the chloroplast translation inhibitor chloramphenicol, which will instantaneously block production of ClpP1$_H$. As can be seen in Fig. 7, both ClpP1$_H$ and ClpP1$_L$ were very stable, and only started to decline after a 26-h incubation. This probably reflects the natural turnover of the complex. No evidence was obtained for a conversion of ClpP1$_H$ to ClpP1$_L$. We conclude that the bulk of accumulated ClpP1$_H$ is stable, and that if ClpP1$_L$ is produced from ClpP1$_H$ (by protein splicing or otherwise), this process is limited to a fraction of ClpP1$_H$, probably during or immediately after the formation of the ClpP/R complex.

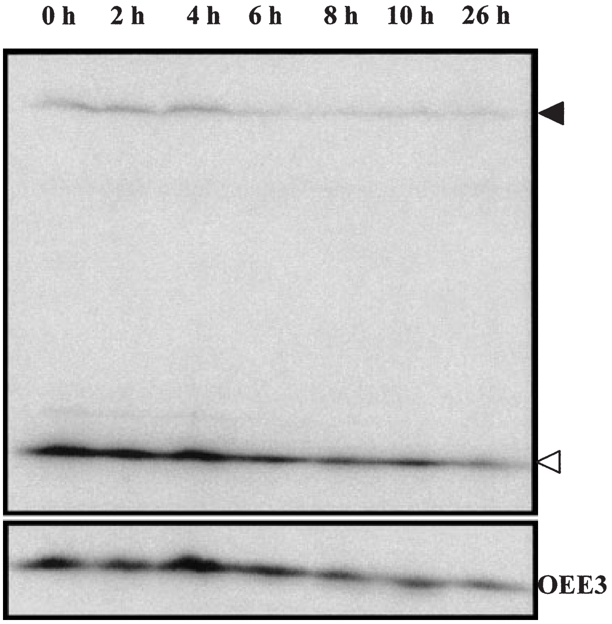


Fig. 7. Chloramphenicol chase experiment. Wild-type cells were treated at t = 0 with chloramphenicol (100 µg·mL^{-1}) to block translation of chloroplast-encoded proteins. Samples collected at various times were separated by electrophoresis and immunoblotted with the ClpP1 antibody. The ClpP1$_H$ and ClpP1$_L$ bands decrease slowly and concomitantly after 24 h, indicating that both are stable in the cell. As a loading control, a duplicate blot was reacted with an antibody to the Photosystem II protein OEE3.

Discussion

The *Chlamydomonas* ClpP complex is substantially larger than its higher plant counterpart

Studies with *Arabidopsis* and other vascular plants have shown that eight nucleus-encoded ClpP/R proteins associate with the plastid-encoded ClpP1 to form an hetero-oligomeric complex which appears identical in composition between various plastid types

[18]. Using 2D electrophoresis, ESI/MS/MS and specific antibodies, we now have identified the components of the chloroplast ClpP complex in the green alga *C. reinhardtii*. We show that it is a hetero-oligomeric assembly of nine ClpP/R proteins, of which seven are nucleus-encoded and two are encoded by the plastid *clpP1* gene. Obviously, some of these subunits must be present in more than one copy, in order to build a tetradecameric complex. We also note that the two halves of the complex cannot be identical, as the total number of gene products is larger than seven.

The most striking difference between the *Chlamydomonas* Clp complex and that of vascular plants is its size (540 kDa vs. 350 kDa), which itself is due at least in part to the presence of the high molecular mass subunit ClpP1$_H$. We had shown before that ClpP1$_H$ is relatively abundant in the cell [25], we now show that it is a stable constitutive subunit of the ClpP/R complex. Because our detection relies entirely on the reaction with ClpP1 and ClpR2 antibodies, we cannot rule out that other ClpP/R complexes lacking ClpP1 and ClpR2 are also present as a smaller complex. However, the narrow band in CN/PAGE (see Figs 3 and 4) does suggest a unique stoichiometry for the *Chlamydomonas* ClpP complex, similarly as in plastids of the *Brassicaceae*. Since one of the subunits, ClpP1$_H$, is two to three times larger than the others, any variation in its stoichiometry is unlikely as it would lead to a heterogeneity in electrophoretic mobility. If no other subunit is present than those reported here, the increase in molecular mass compared to vascular plants must be explained by the presence of the high molecular mass ClpP1$_H$, and to a lesser extend by the larger size of some of the other subunits (ClpP4, ClpR1, ClpR3). Based on these considerations, we propose a stoichiometry of at least three or four copies of ClpP1$_H$ per complex. At this stage, we cannot exclude comigration of different complexes of approximately the same mass, differing only in the stochiometry of some of the lower molecular mass subunits. In plants, there is indirect evidence for variations in the composition of the complex [32].

Structural and functional consequences of the presence of the high Mr ClpP1$_H$ subunit

The presence of ClpP1$_H$ in the complex imposes strong structural constraints on the interaction with chaperones, hence on proteolysis. Assuming that the N- and C-terminal domains of ClpP1$_H$ fold similarly as the corresponding region in other ClpP proteins, then its IS1 domain, inserted between helix 2 and strand 2, must protrude from the apical surface very close to the presumed site of interaction with ClpC. The disrupted loop contains some of the hydrophobic residues that Kim *et al.* proposed to dock the IGF loop in the chaperone [33]. Because of the large size of IS1 (30 kDa), interaction with a Hsp100 chaperone is probably impossible on the apical surface of a ClpP1$_H$-containing heptameric ring. Thus, if the *Chlamydomonas* ClpP complex is to carry out ATP-dependent proteolysis in combination with ClpC, we must hypothesize that ClpP1$_H$ is found only in one of the heptamers, and that only the other one can dock the chaperone.

This must be brought in register with the observation that the *Chlamydomonas* ClpP complex shows no trace of ClpS1 or ClpS2, which in vascular plants are tightly bound subunits of the complex [18]. Extensive search in the *Chlamydomonas* genome and EST databases failed to identify a homolog for these proteins, and similar results were obtained when the *Ostreococcus* and red algal genomes were queried. Thus, the ClpS1 and ClpS2 proteins seem to be restricted to land plants, as suggested before [18]. These proteins, highly similar to the N-terminal domain of HSP100 chaperones, are believed to associate with the apical surface of the complex, making it unable to dock the chaperone. We propose that IS1 in *Chlamydomonas* and *Volvox* plays a similar role, prohibiting access of one side of the ClpP/R complex to the chaperones. Access to the proteolytic chamber of the complex would be possible through only one of its axial pores.

Other models of course are possible, but they seem less likely. For example IS1 itself may be able to bind ClpC, or another chaperone, thus allowing, albeit through a markedly different mechanism, coupling of ATP-dependent protein unfolding with protein degradation. Alternatively, IS1 could by itself carry out the functions normally devoted to the chaperone: substrate recognition and unfolding. Still, it is difficult to reconcile such elaborate functions with the narrow taxonomic distribution of IS1 and its relatively fast evolution rate.

Biogenesis of ClpP1$_L$

Our results show that a polypeptide of ≈25 kDa, which we call ClpP1$_L$, is part of the *Chlamydomonas* ClpP/R complex. Its is recognized by an antibody raised against ClpP1$_H$. Two peptides (totaling 27 residues, including the bordering Arg) were identified by MS in the region of the 2D gel where ClpP1$_L$ migrates, that are absolutely identical to sequences deduced from *clpP1*. No extended similarity exists between ClpP1 and other ClpP/R proteins, so that these peptides must be derived from a product of *clpP1* itself. Extensive searches, including nonassembled reads from the nuclear genome and EST databases failed to identify sequences with high similarity to ClpP1. We conclude that both ClpP1_H and ClpP1_L are products of the chloroplast *clpP1* gene. Experiments are underway to test the protein splicing hypothesis and determine if and how ClpP1_L is generated from ClpP1_H.

Origin and evolution of ClpP/R proteins

Except for the mitochondrial ClpP2, the diversity of plant ClpP/R genes does not seem to correspond to a diversity of peptidases, as originally proposed [34], but to a diversity of structural and catalytic roles within the hetero-oligomeric plastidial ClpP complex. The clear orthology between the *Chlamydomonas* and *Arabidopsis clpP1, CLPP5, CLPR1, CLPR2, CLPR3* and *CLPR4* genes, as well between the algal *CLPR6* and the *CLPP6* of higher plants, indicates that the hetero-oligomeric organization of the ClpP complex was established early in the green lineage, and maintained by strong functional constraints. The only change occurred when the ancestral *CLPP4* duplicated to give *CLPP3* and *CLPP4*.

A general evolutionary trend has been to render more and more of the ClpP subunits inactive (Fig. 1). This occurred first when a ClpR gene appeared in Cyanobacteria (and independently in other bacterial lineages, see supplemental Fig. 1). Then ClpR2 arose, and ClpR6 in green algae. Noncatalytic subunits are also found in the eukaryotic proteasome. Inactive isoforms of FtsH are also found in the *Arabidopsis* genome [4], although the proteins have not been identified yet. It is possible that they participate in hetero-oligomeric complexes similar to those identified for active FtsH isoforms [35].

The diversification of ClpP itself began in Cyanobacteria, with one (*Gleobacter*), two (*Thermosynechococcus*) or three (the general case) isoforms. Our phylogenetic analysis suggests that the cyanobacterial ClpP1 and ClpP2 (*sensu Synechococcus*) have no descendant in plants: large scale alignments including more bacterial sequences (available at the EMBL-Align database as ALIGN_000912) place the origin of the eukaryotic ClpP2 within Proteobacteria rather than Cyanobacteria. This suggests that it was inherited, together with the upstream *clpX* gene, from the ancestral mitochondrial, rather than plastidial, endosymbiont. The third cyanobacterial ClpP (called ClpP3 in *Synechococcus*) is undoubtedly at the origin of ClpP1, which is plastid-encoded in the green lineage and nucleus-encoded in the red lineage and in secondary endosymbionts. All of these sequences start with MPIGVP, in the case of nuclear genes preceded by an N-terminal targeting peptide. In the mitochondrial and bacterial enzymes, the N-terminal end of ClpP has been found to lie in the internal chamber, and the flexible N-terminal loop proposed to play a role in substrate translocation [15,36]. The conservation of the N-terminal sequence in ClpP1 suggests a similar role. A similar sequence is also found in the nuclear-encoded ClpR2 proteins, although less well conserved (Fig. S1). We propose that it also lies at the mature N terminus. Based on the phylogenetic analysis (Fig. 1), ClpR2 proteins certainly derive from the same cyanobacterial ClpP3 ancestor as ClpP1, but have incurred mutations in two or three of the catalytic site residues. The various trees that can be constructed are ambiguous as to whether this duplication occurred prior to the separation of the green and red lineages, or afterwards.

Similarly, the ClpR1, ClpR3 and ClpR4 proteins originate from the cyanobacterial ClpR. This branch is characterized by the presence of a stretch of two to four proline residues just before the first helix and by an eight to nine-residue extension of the loop between strand 2 and helix 3 (L1 loop). This extension has been hypothesized to influence access to the substrate binding pocket [18]. Another interesting feature of these proteins is the mutation into a bulkier residue of the second of two G residues found just before helix 3, which form the top of the substrate binding pocket and are highly conserved in ClpP-type subunits. Thus, this type of ClpR proteins not only have accumulated mutations in active site residues, but also appear to have evolved ways to prevent binding of the substrate in the vicinity of their active site. The gene duplications that gave ClpR1, R3 and R4 are specific to the green lineage, as only one ClpR is found in the red algal, diatom and Plasmodium genomes.

In Cyanobacteria, the ClpP3 and ClpR genes are always found in tandem and cotranscribed. Both are essential in *Synechocystis* sp. PCC 6803 and in *Synechococcus* sp. PCC 7942 [4,37]. In the latter, their protein levels vary in a coordinated fashion during stress or in a ClpP2 deletion strain [37]. This, plus their ancestrality to at least some of the subunits of the plant ClpP/R complex, leads us to propose that ClpP3 and ClpR together form a complex in Cyanobacteria, the ancestor of that in plants. In this view, the simplest evolutionary scenario suggests that ClpP1 and ClpR2 subunits have taken on the positions occupied by ClpP3 in the cyanobacterial complex, while ClpR1, R3 and R4 occupy those of ClpR subunits. As the origin of the remaining ClpP proteins (P3, P4, P5, P6/R6) is unclear, their position in the complex is not predicted by this model. Note that ClpP4 proteins in the red lineage are not strictly speaking orthologous to that in green organisms.

A simpler situation prevails in Apicomplexan parasites such as *Plasmodium* spp. (three genomes) and *Toxoplasma gondii* which contain only one ClpP and one ClpR. As the exact position of the primary endosymbiont is not known, this could reflect either an ancestral organization or secondary gene losses. These proteins are probably targeted to the apicoplast, as *Cryptosporidium parvum*, an Apicomplexan that has lost its apicoplast, also has lost both ClpP and ClpR [38,39]. In any event, apicomplexan ClpP appears as an interesting target for the design of antiparasitic drugs, should it turn out to be essential as its plant homolog is.

In summary, we show that the *Chlamydomonas* ClpP/R complex differs from that in other organisms, in containing several copies of a high molecular mass subunit, derived from an unusually large clpP1 gene. This protein (clpP1$_H$) is predicted to expose a large IS1 domain on the apical surface of the ClpP barrel, probably interfering with the docking of HSP100 chaperones.

Experimental procedures

Database searches and sequence alignment

The cDNA sequences of *Chlamydomonas* nuclear *CLPP/R* genes were collected from the *Chlamydomonas* EST project (www.chlamy.org) or determined through assembly by CAP3 [40] of ESTs and additional internal sequences obtained from clones of the Kazusa EST project (www.kazusa.or.jp/en/plant/chlamy/EST/). *Chlamydomonas* genomic sequences and *V. carteri* whole genome shotgun sequences were collected from the JGI website (http://genome.jgi-psf.org/chlre2/chlre2.home.html and http://genome.jgi-psf.org/chlre1/chlre1.home.html). *Ostreococcus tauri* sequence contigs were searched by BLAST courtesy of Hervé Moreau (UMR 7628, Banuyls, France).

Protein sequences were aligned with CLUSTALW, and the alignment was edited with the program BIOEDIT. Protein distances were calculated using the PAM matrix, and a Neighbor phylogenic tree (randomized: 65; 5) was derived with the PHYLIP package and visualized with TREEVIEW. Predictions for chloroplast localization were made using TARGETP [41] and PREDOTAR [42].

MS

Stained protein spots were excised from the gel, washed, reduced with dithiothreitol, alkylated with iodoacetamide, and digested with modified trypsin (Promega, Charbonnières, France) as described in [43]. The peptides were extracted and dissolved in 20 µL 5% formic acid and applied to the MALDI-TOF target plate by the dried droplet method using α-cyano-4-hydrocynnamic acid as matrix. When necessary, the samples were concentrated using microcolumns [44] and eluted directly onto the MALDI target. The mass spectra were obtained using a MALDI-TOF mass spectrometer (Voyager-DE-STR, Perseptive Biosystems Inc., Framingham, MA). The spectra were annotated with the program M/Z from Proteometrics and internally calibrated using tryptic peptides from autodigestion. The resulting peptide mass lists were searched against the latest version of the NCBI nonredundant database using the search engine ProFound and a database of Clp proteins from *Chlamydomonas* made in-house, using the PS1 solution engine from Perseptive Biosystems. The search strategy was in principle as described previously [43]. To analyze the samples further, the remainder of the extracted peptides were desalted and concentrated on microcolumns (Poros R2, PE Biosystems, Foster City, CA) and eluted directly into nanoelectrospray needles (Protana A/S, Odense, Denmark) with 1.2 µL 50% MeOH and 1% formic acid [45]. The spectra were acquired on an electrospray tandem mass spectrometer (Q-TOF, Micromass, Millford, MA). The instrument was calibrated with 1 $\mu g \cdot \mu L^{-1}$ NaI in 50% isopropanol. The spectra were used to search the public databases with the program MASCOT. Alternatively, the MS/MS spectra were interpreted using MASSLYNX and PEPSEQ (Micromass) and were used to search different public databases and the in-house' Clp database using FASTA3.

Growth of *Chlamydomonas* and subcellular fractionation

Chlamydomonas reinhardtii wild-type and cell-wall less mutant (*cw15*) strains were grown on Tris-Acetate medium [46] at 25 °C under 30 $\mu E \cdot m^{-2}\, s^{-1}$ continuous illumination. The *rbcL* mutant (*rbcL-18-5b*, kindly provided by R.L. Spreitzer) and a double mutant *cw15 rbcL-18-5b* (kindly provided by Katia Wostrikoff) were grown in the dark. Wild-type cells were broken by a single passage through a French press operated at 6000 psi. The soluble (stroma) fraction was recovered after centrifugation at 100 000 ***g*** for 20 min. Chloroplasts were purified from cell wall-less strains essentially as in [47], washed twice in medium I and broken by resuspension in 10 mM Hepes/KOH pH 7.5, 5 mM $MgCl_2$ and repeated passage through a fine gauge needle. Membranes were removed by ultracentrifugation (20 min, 100 000 ***g***). In some experiments, chloroplasts or membrane fractions were treated by passage through a Yeda press (Yeda, Rehovot, Israel) at 10^7 Pa.

Denaturing electrophoresis

Denaturing electrophoresis (SDS/PAGE) was carried out in 7.5–15% acrylamide gradient gels, or in 12–18% gels containing 8 M urea [48]. Proteins were electroblotted onto nitrocellulose membrane [49]. Immunoblots were revealed either with [^{125}I]-labeled protein-A [50] and scanned using a phosphorimager scanner (Molecular Dynamics, Sunnyvale, CA), or with the ECL system (Amersham, Louisville, CO). The ClpR2 antiserum was raised against the peptide KMPSTGPSFKFERQNDE, corresponding to residues 214–236, after C-terminal coupling to keyhole-limpet hemocyanin. The anti-ClpP1 serum has been raised against the entire ORF of *C. reinhardtii clpP1* [25].

Colorless-native PAGE

Chloroplast soluble protein fractions were supplemented with 500 mM amino-6-caproic acid, 50 mM Bistris HCl (pH 7.4), 15% glycerol and 0.004% Ponceau Red. Protein extracts were loaded onto a 0.75-mm thick, 18-cm long polyacrylamide gradient gel [51]. Electrophoresis was carried out overnight at 4 °C, at 350 V. Molecular weight markers were BSA (60 kDa, 120 kDa), bovine liver catalase (220 kDa), *E. coli* β-galactosidase (465 kDa) and bovine thyroglobulin (670 kDa). First dimension gel lanes were cut out and equilibrated in 1% SDS, 10 mM dithiothreitol, 2 mM EDTA and 100 mM Na_2CO_3 for 15 min at 65 °C. Gel strips were then washed five times at room temperature in Laemmli/SDS buffer. Gel slices were loaded on a second dimension 1.5-mm thick 12–18% polyacrylamide/urea gels, immobilized with Laemmli stacking gel and migrated overnight. Gels were stained using Coomassie blue G-250 (Biosafe, Biorad, Marnes-la-Coquette, France) or transferred onto a nitrocellulose membrane (Hybond-ECL, Amersham) for immunoblotting.

Acknowledgements

This work was supported by the CNRS UPR1261, by a grant from the Novartis Foundation to W.M and by the National Science Foundation – MCB #343444- to KJvW. We thank Katia Wostrikoff for the construction of the *cw15; rbcl-18–5b* double mutant. The *Chlamydomonas, Volvox* and *Thalassiosira* genome sequence data appear courtesy of the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/). We thank the Kazusa Institute and the Chlamydomonas Genome Project for EST data and cDNA material, and H. Moreau (Observatoire Océanologique de Banyuls, France) for granting us access to the *Ostreococcus* BLAST analysis.

References

1 Lupas A, Flanagan JM, Tamura T & Baumeister W (1997) Self-compartmentalizing proteases, *Trends Biochem Sci.* **22**, 99–404.

2 Adam Z, Adamska I, Nakabayashi K, Ostersetzer O, Haussuhl K, Manuell A, Zheng B, Vallon O, Rodermel SR, Shinozaki K & Clarke AK (2001) Chloroplast and mitochondrial proteases in arabidopsis. a proposed nomenclature. *Plant Physiol* **125**, 1912–1918.

3 Gottesman S, Squires C, Pichersky E, Carrington M, Hobbs M, Mattick JS, Dalrymple B, Kuramitsu H, Shiroza T & Foster T (1990) Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. *Proc Natl Acad Sci USA* **87**, 3513–3517.

4 Sokolenko A, Pojidaeva E, Zinchenko V, Panichkin V, Glaser VM, Herrmann RG & Shestakov SV (2002) The gene complement for proteolysis in the cyanobacterium *Synechocystis* sp. PCC 6803 and *Arabidopsis thaliana* chloroplasts. *Curr Genet* **41**, 291–310.

5 Woo KM, Chung WJ, Ha DB, Goldberg AL & Chung CH (1989) Protease Ti from *Escherichia coli* requires ATP hydrolysis for protein breakdown but not for hydrolysis of small peptides. *J Biol Chem* **264**, 2088–2091.

6 Maurizi MR, Thompson MW, Singh SK & Kim SH (1994) Endopeptidase Clp: ATP-dependent Clp protease from *Escherichia coli. Methods Enzymol* **244**, 314–331.

7 Grimaud R, Kessel M, Beuron F, Steven AC & Maurizi MR (1998) Enzymatic and structural similarities between the *Escherichia coli* ATP-dependent proteases, ClpXP and ClpAP. *J Biol Chem* **273**, 12476–12481.

8 Maurizi MR, Clark WP, Katayama Y, Rudikoff S, Pumphrey J, Bowers B & Gottesman S (1990) Sequence and structure of Clp P, the proteolytic component of the ATP- dependent Clp protease of *Escherichia coli. J Biol Chem* **265**, 12536–12545.

9 Wang J, Hartling JA & Flanagan JM (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**, 447–456.

10 Wojtkowiak D, Georgopoulos C & Zylicz M (1993) Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of *Escherichia coli. J Biol Chem* **268**, 22609–22617.

11 Gottesman S, Clark WP & Maurizi MR (1990) The ATP-dependent Clp protease of *Escherichia coli.* Sequence of clpA and identification of a Clp-specific substrate. *J Biol Chem* **265**, 7886–7893.

12 Flynn JM, Neher SB, Kim YI, Sauer RT & Baker TA (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**, 671–683.

13 Levchenko I, Seidel M, Sauer RT & Baker TA (2000) A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**, 2354–2356.

14 Guo F, Esser L, Singh SK, Maurizi MR & Xia D (2002) Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *J Biol Chem* **277**, 46753–46762.

15 Kang SG, Maurizi MR, Thompson M, Mueser T & Ahvazi B (2004) Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP. *J Struct Biol* **148**, 338–352.

16 Halperin T, Zheng B, Itzhaki H, Clarke AK & Adam Z (2001) Plant mitochondria contain proteolytic and regulatory subunits of the ATP-dependent Clp protease. *Plant Mol Biol* **45**, 461–468.

17 Kang SG, Ortega J, Singh SK, Wang N, Huang NN, Steven AC & Maurizi MR (2002) Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP. *J Biol Chem* **277**, 21095–21102.

18 Peltier JB, Ripoll DR, Friso G, Rudella A, Cai Y, Ytterberg J, Giacomelli L, Pillardy J & van Wijk KJ (2004) Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. *J Biol Chem* **279**, 4768–4781.

19 Halperin T, Ostersetzer O & Adam Z (2001) ATP–dependent association between subunits of Clp protease in pea chloroplasts. *Planta* **213**, 614–619.

20 Weaver LM, Froehlich JE & Amasino RM (1999) Chloroplast-targeted ERD1 protein declines but its mRNA increases during senescence in *Arabidopsis*. *Plant Physiol* **119**, 1209–1216.

21 Peltier JB, Ytterberg J, Liberles DA, Roepstorff P & van Wijk KJ (2001) Identification of a 350 kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *J Biol Chem* **276**, 16318–16327.

22 Clarke AK, Schelin J & Porankiewicz J (1998) Inactivation of the clpP1 gene for the proteolytic subunit of the ATP-dependent Clp protease in the cyanobacterium *Synechococcus* limits growth and light acclimation. *Plant Mol Biol* **37**, 791–801.

23 Huang C, Wang S, Chen L, Lemieux C, Otis C, Turmel M & Liu XQ (1994) The *Chlamydomonas* chloroplast clpP gene contains translated large insertion sequences and is essential for cell growth. *Mol Gen Genet* **244**, 151–159.

24 Shikanai T, Shimizu K, Ueda K, Nishimura Y, Kuroiwa T & Hashimoto T (2001) The chloroplast clpP gene, encoding a proteolytic subunit of ATP- dependent protease, is indispensable for chloroplast development in tobacco. *Plant Cell Physiol* **42**, 264–273.

25 Majeran W, Wollman FA & Vallon O (2000) Evidence for a role of ClpP in the degradation of the chloroplast cytochrome b_6f complex. *Plant Cell* **12**, 137–150.

26 Majeran W, Olive J, Drapier D, Vallon O & Wollman FA (2001) The light sensitivity of ATP synthase mutants of *Chlamydomonas reinhardtii*. *Plant Physiol* **126**, 421–433.

27 Wang S & Liu XQ (1997) Identification of an unusual intein in chloroplast ClpP protease of *Chlamydomonas* eugametos. *J Biol Chem* **272**, 11869–11873.

28 Hodges RA, Perler FB, Noren CJ & Jack WE (1992) Protein splicing removes intervening sequences in an archaea DNA polymerase. *Nucl Acids Res* **20**, 6153–6157.

29 Paulus H (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* **69**, 447–496.

30 Frishman D & Argos P (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng* **9**, 133–142.

31 Garnier J, Gibrat JF & Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol* **266**, 540–553.

32 Sjogren LL, Macdonald TM, Sutinen S & Clarke AK (2004) Inactivation of the clpC1 gene encoding a chloroplast Hsp100 molecular chaperone causes growth retardation, leaf chlorosis, lower photosynthetic activity, and a specific reduction in photosystem content. *Plant Physiol* **136**, 4114–4126.

33 Kim YI, Levchenko I, Fraczkowska K, Woodruff RV, Sauer RT & Baker TA (2001) Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat Struct Biol* **8**, 230–233.

34 Porankiewicz J, Wang J & Clarke AK (1999) New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. *Mol Microbiol* **32**, 449–458.

35 YuF, Park S & Rodermel SR (2004) The Arabidopsis FtsH metalloprotease gene family: interchangeability of subunits in chloroplast oligomeric complexes. *Plant J* **37**, 864–876.

36 Gribun A, Kimber MS, Ching R, Sprangers R, Fiebig KM & Houry WA (2005) The ClpP double-ring tetradecameric protease exhibits plastic ring–ring interactions and the N-termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *J Biol Chem* **280**, 16185–16196.

37 Schelin J, Lindmark F & Clarke AK (2002) The clpP multigene family for the ATP-dependent Clp protease in the cyanobacterium *Synechococcus*. *Microbiology* **148**, 2255–2265.

38 Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L & Kapur V (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445.

39 Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS & Kissinger JC (2004) Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol* **5**, R88.

40 Huang X & Madan A (1999) CAP3: a DNA Sequence Assembly program. *Genome Res* **9**, 868–877.

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

42 Small I, Peeters N, Legeai F & Lurin C (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581–1590.

43 Shevchenko A, Wilm M, Vorm O & Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850–858.

44 Gobom J, Nordhoff E, Mirgorodskaya E, Ekman R & Roepstorff P (1999) Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J Mass Spectrom* **34**, 105–116.

45 Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T & Mann M (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469.

46 Harris EH (1989) *The Chlamydomonas Source Book: a Comprehensive Guide to Biology and Laboratory Use*. Academic Press, San Diego.

47 Schroda M, Vallon O, Whitelegge JP, Beck CF & Wollman FA (2001) The chloroplastic GrpE homolog of Chlamydomonas: two isoforms generated by differential splicing. *Plant Cell* **13**, 2823–2839.

48 Piccioni RG, Bennoun P & Chua NH (1981) A nuclear mutant of Chlamydomonas reinhardtii defective in photosynthetic photophosphorylation. Characterization of the algal coupling factor ATPase. *Eur J Biochem* **117**, 93–102.

49 Towbin H, Staehelin T & Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* **76**, 4350–4354.

50 Burnette WN (1981) 'Western blotting': electrophoretic transfer of proteins from sodium dodecyl sulfate–polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* **112**, 195–203.

51 Schagger H, Cramer WA & von Jagow G (1994) Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* **217**, 220–230.

Supplementary material

The following supplementary material is available for this article online:

Figure S1. Alignment of ClpP/R proteins, generated by CLUSTALW and edited manually. It includes proteins from photosynthetic and nonphotosynthetic eukaryotes and eubacteria, as depicted in the table. Sequences have been trimmed at the N and C termini to leave only the ClpP domain. Top: structure of *E. coli* ClpP (1TYF). Residues involved in catalysis are marked by *; those involved in binding the IGF loop of the chaperone are marked by ! (hydrophobic) or X (the D residue). The location of IS1 in *Chlamydomonas* is marked by an arrow. The loop insertion in ClpR proteins is boxed.