

Multistep Processing of an Insertion Sequence in an Essential Subunit of the Chloroplast ClpP Complex^{*[S]}

Received for publication, December 24, 2008, and in revised form, April 2, 2009. Published, JBC Papers in Press, April 3, 2009, DOI 10.1074/jbc.M109.002733

Benoit Derrien^{†1}, Wojciech Majeran[§], Francis-André Wollman[‡], and Olivier Vallon^{‡2}

From the [‡]Institut de Biologie Physico-Chimique, UMR7141 CNRS/Université Pierre et Marie Curie, 75005 Paris, France and the [§]Department of Plant Biology, Cornell University, Ithaca, New York 14853

In *Chlamydomonas reinhardtii*, the *clpP1* chloroplast gene encoding one of the catalytic subunits of the ClpP protease complex contains a large in-frame insertion sequence (IS1). Based on the *Escherichia coli* ClpP structure, IS1 is predicted to protrude at the apical surface of the complex, likely influencing the interaction of the catalytic core with ClpC/HSP100 chaperones. Immunoblotting with an anti-ClpP1 antibody detected two immunoreactive forms of ClpP1: ClpP1_H (59 kDa) and ClpP1_L (25 kDa). It has been proposed that IS1 is a new type of protein intron (different from inteins). By studying transformants harboring mutations at the predicted borders of IS1 and tags at the C terminus of ClpP1 (tandem affinity purification tag, His tag, Strep-Tag) or within the IS1 sequence (3-hemagglutinin tag), we show that IS1 is not a protein intron and that ClpP1_L results from endoproteolytic cleavage inside IS1. Processing sites have been identified in the middle of IS1 and near its C terminus. The sites can be mutated without abolishing processing.

Clp proteases are self-compartmentalized serine proteases present in most eubacteria and, as a consequence of endosymbiotic events, in the mitochondrion and chloroplast of eukaryotes. In *Escherichia coli*, the organism in which they have been best characterized, Clp proteases associate a homo-oligomeric peptidase (ClpP) and a chaperone (ClpA or ClpX) that belongs to the Clp/HSP100 family, itself part of the large group of AAA⁺ ATPases (1–4). ClpP is composed of 14 identical subunits arranged in two heptameric rings related by central symmetry. They form a barrel-like structure with the 14 active sites facing an inner proteolytic chamber (5). ClpP alone is able to degrade only small peptides (6), and the recognition and unfolding of protein substrates are carried out by the Clp/HSP100 chaperone. The chaperone docks on the apical surfaces of ClpP and uses ATP hydrolysis to unfold and feed substrates through the ClpP axial pore into the proteolytic chamber (7–10).

In chloroplasts, ClpP is present as a hetero-oligomer associating up to eight different types of subunit. This is the result of a gene diversification process that has begun in cyanobacteria

and continues in the chloroplast of land plants. Not only has the number of *clpP* genes grown, but *clpR* genes have appeared that carry mutations in at least one residue of the catalytic triad and are thus presumed catalytically inactive. In the green alga *Chlamydomonas reinhardtii*, three *clpP* genes (*clpP1*, *CLPP4*, and *CLPP5*) and five *clpR* genes (*CLPR1*–*CLPR4* and *CLPR6*) code for the subunits of the chloroplast ClpP complex (11). An additional *CLPP2* gene codes for the homo-oligomeric mitochondrial ClpP.

ClpP1 is the only subunit that is encoded in the chloroplast and probably the best conserved. In *C. reinhardtii*, *clpP1* contains a large insertion sequence (IS1)³ translated in-frame with the conserved N- and C-terminal regions. This results in a protein about twice as large (~59 kDa) as in other organisms. *Chlamydomonas* ClpP1 can be divided into two sequence domains, SD1 and SD2 (the latter containing the catalytic residues), corresponding to the conserved sequence, and one insertion sequence, IS1 (12). In *C. reinhardtii*, antisera raised against the entire open reading frame (ORF) recognize two products of *clpP1* in Western blot: ClpP1_H (59 kDa) and ClpP1_L (21 kDa) (13). As the *clpP1* mRNA does not undergo splicing (12), it has been proposed that IS1 could be a protein intron. Protein introns such as inteins (14) are defined as in-frame intervening sequences that disrupt a host gene and are post-translationally excised by a self-catalytic mechanism. In the case of *clpP1*, ClpP1_H would be the precursor protein and ClpP1_L the spliced form. However, IS1 lacks the sequence motifs characteristic of inteins. In addition, both ClpP1_L and ClpP1_H are stable, and both associate in the 540-kDa ClpP complex (11). Thus, if IS1 were a protein intron, it would be an unusual type. In the related species *Chlamydomonas eugametos*, *clpP1* contains, in addition to IS1, another insertion sequence (IS2) displaying most of the sequence features of inteins. Indeed, IS2 can be induced to self-splice in *E. coli* by changing a single residue (15).

In this study, we show that IS1 is not a protein intron and that ClpP1_L is the product of a complex proteolytic maturation of ClpP1_H. We have found similar insertion sequences in the *clpP1* genes of other green algae from the group Chlorophyceae. Green algae accumulate such insertion sequences in many of their chloroplast genes, probably as a result of a high frequency of genome rearrangements.

* This work was supported by CNRS UMR7141 and the Université Pierre et Marie Curie-Paris 6.

[S] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental "Experimental Procedures," Figs. S1–S5, and Table I.

¹ Adjoint Temporaire d'Enseignement et de Recherche at the Université Pierre et Marie Curie-Paris 6.

² To whom correspondence should be addressed: IBPC, UMR7141 CNRS/UPMC, 13 rue Pierre et Marie Curie, 75005 Paris, France. Tel.: 33-1-5841-5058; Fax: 33-1-5841-5022; E-mail: ovallon@ibpc.fr.

³ The abbreviations used are: IS, insertion sequence; SD, sequence domain; ORF, open reading frame; WT, wild-type; TAP, tandem affinity purification; HA, hemagglutinin.

EXPERIMENTAL PROCEDURES

Strains and Culture Conditions—Our *C. reinhardtii* wild-type (WT) strain WTS34 *mt*⁺ is derived from the 137c background. Other algal strains were kindly provided by Pierre Cardol (Liège University, Liège, Belgium). All strains were grown in Tris acetate/phosphate medium under constant illumination (5–10 microeinsteins m⁻² s⁻¹) at 25 °C (16).

ClpP1 Mutagenesis—A genomic fragment corresponding to nucleotides 16972–20083 of the chloroplast genome (GenBank™ accession number BK000554) and encompassing the C-terminal end of the *petB* coding region and the *trnL*, *clpP1*, *trnW*, and *trnS* genes was PCR-amplified using *Chlamydomonas* total DNA as template and oligonucleotide primers SBamHI and REVA (see supplemental “Experimental Procedures” for oligonucleotide sequences and details on site-directed mutagenesis). After digestion with BamHI and KpnI, it was cloned into the pBluescript KS(–) vector digested by the same enzymes to create plasmid pClpP1. To create pLG3S, the 1.9-kilobase pair *aadA* cassette conferring resistance to spectinomycin and streptomycin (17) was inserted in the unique EcoRV site so that *aadA* and *clpP1* read in the same direction. Site-directed mutagenesis was carried out with PCR or megaprimer PCR (18). We verified correct introduction of all mutations by partial sequencing of the plasmids.

Chloroplast Transformation—Chloroplast transformation by tungsten particle bombardment (19) was performed with a helium gun built in the laboratory by D. Béal. Cells were grown in liquid Tris acetate/phosphate medium to a density of 2 × 10⁶ cells ml⁻¹. 10⁸ cells were plated on Tris acetate/phosphate medium supplemented with 100 μg ml⁻¹ spectinomycin and bombarded with tungsten particles (diameter, ~1 μm) coated with the appropriate DNA. Colonies were picked up after 2 weeks of growth and subcloned on Tris acetate/phosphate plates containing increasing concentrations of spectinomycin (from 100 to 500 μg ml⁻¹) until they reached homoplasmy (20). The presence of transforming DNA was assessed by PCR with specific primers and restriction analysis of the PCR products, whereas that of WT DNA was assessed by PCR using primers Seq4 and WA2. Colonies recovered from plates were resuspended in 20 μl of sterile water, out of which 10 μl were used for Chelex 100 DNA extraction (21), whereas the remaining 10 μl were used for the next subcloning round.

Biochemical Analysis—SDS-PAGE was performed on 12–18% acrylamide gels containing 8 M urea (22); cell samples were loaded on an equal chlorophyll basis. For purified complex, protein concentration was determined using a BCA assay kit (Sigma). Proteins were electroblotted onto nitrocellulose membranes (Hybond-ECL, GE Healthcare) (23). Immunodetection was performed by ECL (GE Healthcare). The ClpP1 antiserum was raised against the entire ORF of *C. reinhardtii clpP1* (13). The IS1 antiserum was raised by Neosystem (Strasbourg, France) against peptide NNESGRSLYRKQTER, corresponding to residues 325–339 localized in the C-terminal part of IS1. For Edman sequencing of ClpP1 fragments, proteins were transferred onto Westran polyvinylidene difluoride (Whatman) and stained with Coomassie Blue R-250. Edman

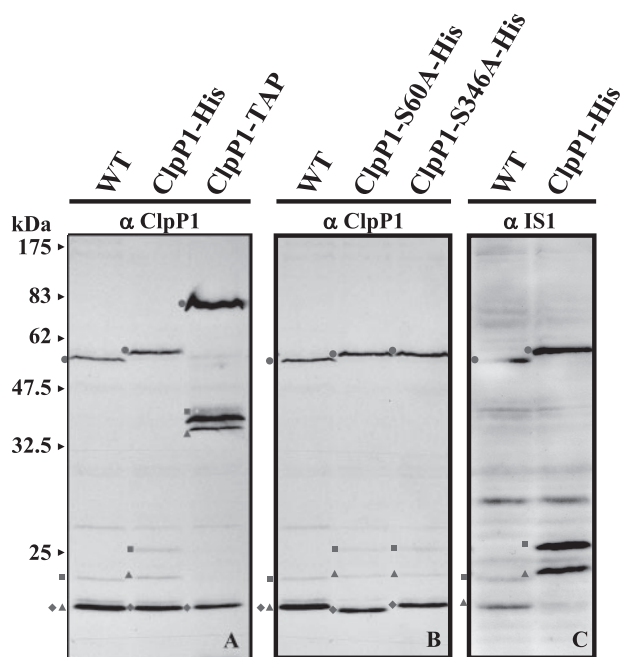


FIGURE 1. Immunoblot analysis of C-terminally tagged strains (ClpP1-His and ClpP1-TAP) and IS1 border mutants (S60A and S346A). Western blots reacted with antisera raised against ClpP1 (A and B) and an IS1 peptide epitope (C). The products of the *clpP1* gene are indicated: ClpP1_H (●), ClpP1_N (◆), ClpP1_C (■), and ClpP1_{C'} (▲).

sequencing was performed at the Plateau Technique d'Analyse et de Microséquençage des Protéines (Institut Pasteur).

RESULTS

ClpP1 IS1 Is Not a Protein Intron—To test the hypothesis that IS1 is a protein intron, we introduced at the C terminus of ClpP1 a His₁₀ tag, a tandem affinity purification (TAP) tag (Fig. 1A), or a Strep-Tag (see Fig. 4). These tags consist of 10 consecutive histidines; a combination of a calmodulin-binding domain, a tobacco etch virus protease cleavage site, and a protein A domain (24); and a 9-amino acid peptide with high affinity for Strep-Tactin, respectively. In all cases, we observed the expected change in the electrophoretic migration of ClpP1_H (59 kDa in the WT), confirming that this immunoreactive band contains the C terminus and is produced by translation of the complete *clpP1* ORF. In contrast, none of the tagged strains showed a change in the migration of ClpP1_L (~21 kDa in all strains), indicating that ClpP1_L does not contain the C terminus of ClpP1_H, as would be expected if it were produced by IS1 splicing.

Protein introns like inteins rely on their N- and C-terminal flanking residues (usually Cys but sometimes Ser) for excision (25). To further rule out that an intein-like splicing of IS1 produced ClpP1_L, we mutated its putative flanking residues (Ser⁶⁰ and Ser³⁴⁶) to Ala. In the two single mutants produced in the His-tagged context (S60A-His and S346A-His), we observed no modification of the ratio between the ClpP1_H and ClpP1_L bands compared with the WT (Fig. 1B). This result rules out that IS1 is a protein intron, as we would have expected at least a change in splicing efficiency by mutating its bordering residues. Interestingly, the only difference was the slightly faster migration of ClpP1_L caused by mutating Ser⁶⁰ to Ala (Fig. 1B). This was also

Proteolytic Maturation of Chloroplast ClpP1 Subunit

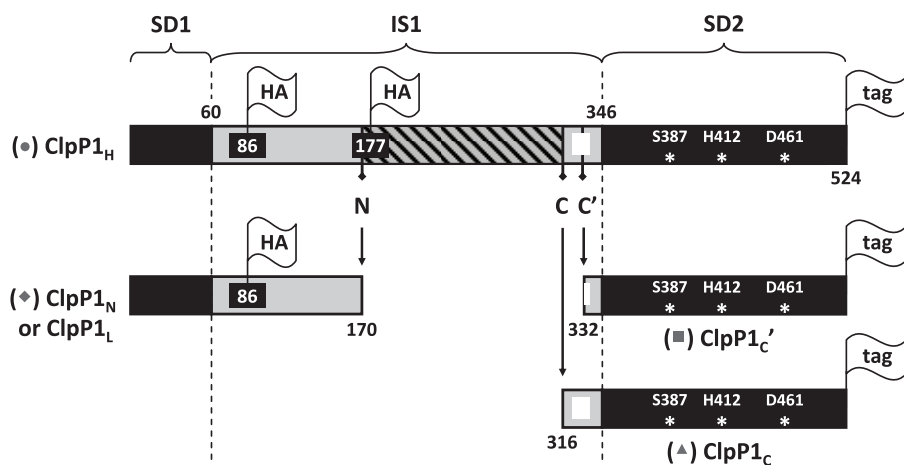


FIGURE 2. Model for the maturation of ClpP1. SD1 and SD2 are colored in black, and IS1 in light gray. The excised portion of IS1 is shown as a cross-hatched box. The positions of the various tags used in this study are indicated by flags. The peptide used to raise the anti-IS1 antibody (N³²⁵NESGRSLYRKQTER³³⁹) is shown as a white box.

observed two weakly immunoreactive bands that could correspond to such C-terminal fragments, with 1.5 kDa added from the His tag. They were not stronger than other background bands, but were missing in the WT. Instead, the WT showed a faint band at ~23 kDa (Fig. 1, ■), which was absent in the tagged strains. These observations led us to propose that ClpP1 is processed to yield (i) an N-terminal fragment, hitherto called ClpP1_L and which we now rename ClpP1_N, and (ii) two C-terminal fragments, which we call ClpP1_C (Fig. 1, ■) and ClpP1_{C'} (Fig. 1, ▲), whose migration is retarded in strains carrying C-terminal tags (Fig. 2). In the WT strain, ClpP1_{C'}

co-migrated with ClpP1_N and was therefore not seen because of the much stronger immunoreactivity of ClpP1_N. It was only revealed when the presence of an epitope tag affected its mobility (Fig. 1) or that of ClpP1_N (see ClpP1HA₈₅ in Fig. 3A below).

Further evidence for this model was obtained by Western blotting with an antibody directed against a peptide epitope (N³²⁵NESGRSLYRKQTER³³⁹) corresponding to a region near the C terminus of IS1 (Fig. 1C). Although not very specific, this antibody recognized in Western blots all the bands that we propose represent C-terminal fragments of ClpP1 (Fig. 1C, ■ and ▲). This result also allowed us to conclude that the N termini of these C-terminal fragments lie upstream of the peptide epitope, *i.e.* within IS1. We consistently observed two closely spaced C-terminal fragments regardless of the tag used, suggesting that cleavage can occur at either of two positions in the C-terminal portion of IS1. We call these cleavage sites C and C' (Fig. 2).

To prove that the smaller fragment (ClpP1_N) corresponds to the N-terminal part of ClpP1, we introduced a 3-hemagglutinin (HA) tag at various positions inside the protein. In strain ClpP1HA₈₅, the tag was inserted at position 85 (Fig. 2), *i.e.* in the N-terminal part of IS1. In this strain, not only the ClpP1_H but also the ClpP1_N bands were shifted upwards by ~4.3 kDa, the size of the tag. Both bands were recognized by the anti-HA antibody (Fig. 3A). In contrast, in strain ClpP1HA₁₇₆, in which the tag was introduced at position 176, approximately in the middle of IS1 (Fig. 2), the migration of ClpP1_N was unchanged. In this strain, the anti-HA antibody labeled ClpP1_H, but not ClpP1_N or the ClpP1_C and ClpP1_{C'} C-terminal bands (Fig. 3B). This indicates that the central portion of IS1, including position 176, is lost during processing. Processing must therefore include another cleavage event at a site between positions 85 and 176, which we call cleavage site N. This is in line with the observation that the sum of the apparent molecular masses of ClpP1_N and ClpP1_C (21 and 23 kDa, respectively, in the WT strain) is less than the mass of the ClpP1_H precursor (59 kDa).

Identification and Mutagenesis of the Cleavage Sites in IS1—To identify the processing sites more precisely, we analyzed the ClpP1-derived bands by N-terminal sequencing (Edman degra-

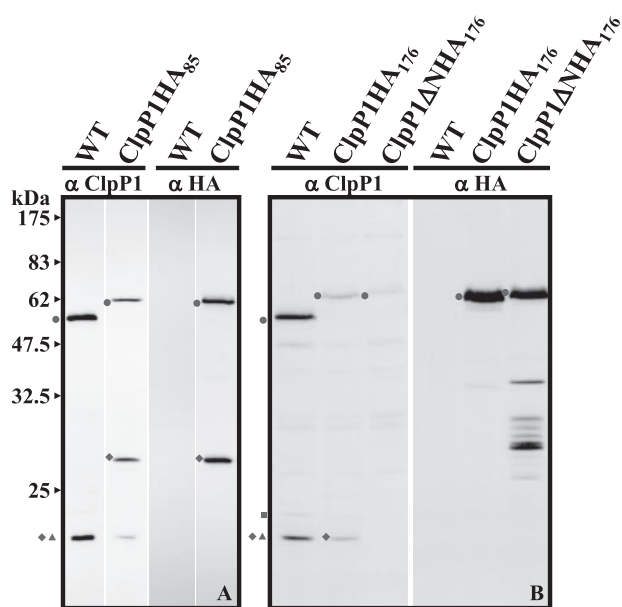


FIGURE 3. Insertion of a 3-HA tag in IS1. Immunoblots reacted with antibodies to ClpP1 and the HA tag. The symbols are the same as indicated in the legend to Fig. 1. A, analysis of ClpP1HA₈₅; B, analysis of ClpP1HA₁₇₆ and ClpP1ΔNHA₁₇₆.

observed in a double mutant (S60A/S346A) generated in a non-tagged background (data not shown). This suggests that ClpP1_L contains this residue and is hence derived from the N-terminal part of the protein.

ClpP1 Is Cleaved within IS1 to Generate N- and C-terminal Fragments—In immunoblot experiments, the TAP-tagged strain showed a very strong labeling of the tagged ClpP1_H band (Fig. 1A, ●, 79 kDa *versus* 59 kDa in the WT) due to the fact that the protein A domain has by itself the capacity to nonspecifically bind antibodies in Western blots. Interestingly, two other strongly labeled bands were visible at 43 and 41 kDa (Fig. 1A, ■ and ▲), which we reasoned should also contain the C terminus of ClpP1. We calculated that these two bands comprised the C-terminal 23 and 21 kDa of the ClpP1 protein, respectively, plus the 20-kDa TAP tag. In the His-tagged strain, we also

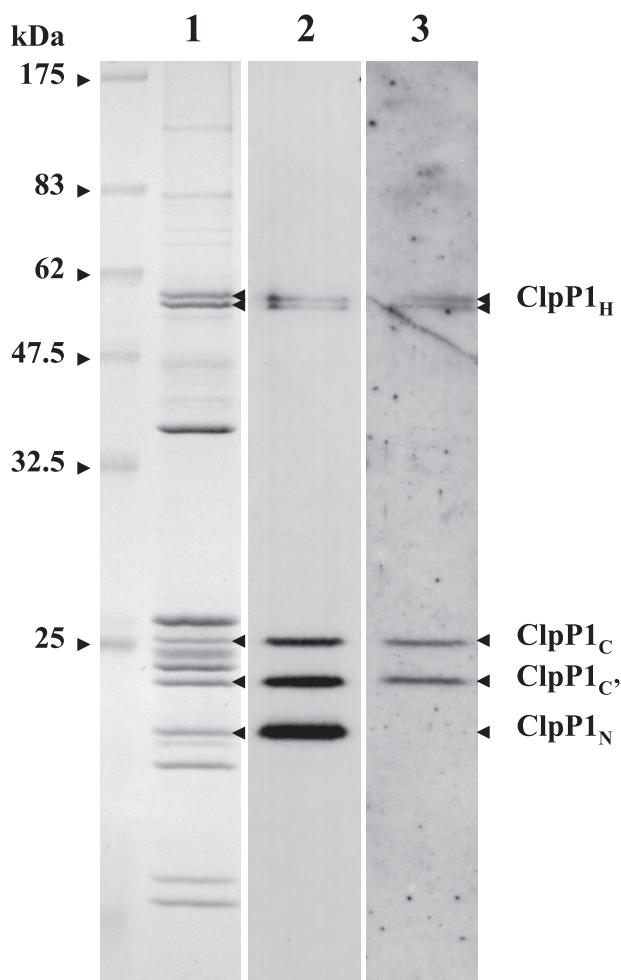


FIGURE 4. Purified Strep-Tag-tagged ClpP complex. Lane 1, Coomassie Blue; lanes 2 and 3, immunoblots probed with antibodies to ClpP1 and the IS1 peptide, respectively.

dation). We used a partially purified preparation of the ClpP complex obtained from the Strep-Tag-tagged strain by affinity chromatography.⁴ This preparation contains all the ClpP and ClpR subunits and the four products of ClpP1, easily identified by immunoblotting (Fig. 4). As expected, the sequence obtained from the ClpP1_N band (XIGV...) matched that expected from the N terminus of the ClpP1 ORF (MPIGV) after removal of the N-terminal Met. The same sequence was obtained from the two bands forming the ClpP1_H doublet. For ClpP1_C and ClpP1_{C'}, two different N-terminal sequences were obtained (NYLD and YRK, respectively), identifying the positions of cleavage sites C and C', respectively (Figs. 2 and 4). They are both found near the C terminus of IS1, and site C' is located within the peptide used to generate the IS1-specific antibody (NNESGRSL↓YRK-QTER). Interestingly, electrospray ionization-tandem mass spectrometry experiments on the purified complex (data not shown) also identified a semitryptic peptide (NYLDQGALN-NESGR) whose N terminus coincided with site C. What could be the role of these processing events? To address this question, we mutated residues around sites C and C', changing the

⁴ B. Derrien, W. Majeran, K.-J. van Wijk, F.-A. Wollman, and O. Vallon, unpublished data.

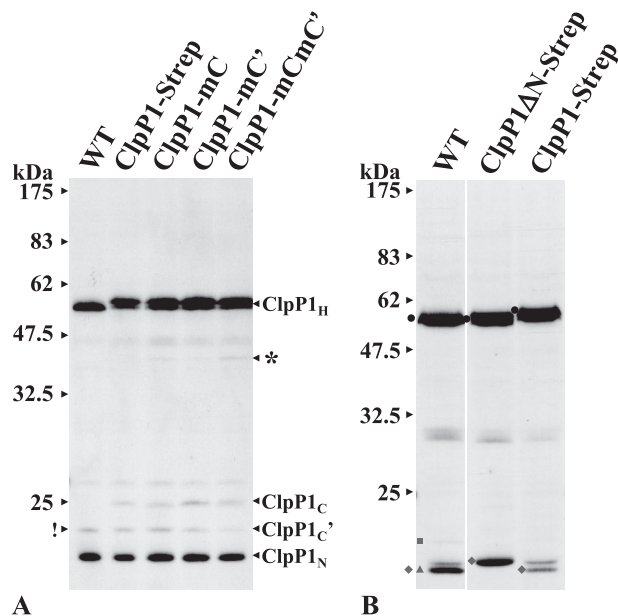


FIGURE 5. Immunoblot analysis of cleavage site C, C', and N mutants. Immunoblots reacted with antibodies to ClpP1. A, site C and C' mutants. Note the new band at ~40 kDa (*). In the WT lane, the band indicated by an exclamation point is not ClpP1_{C'} but ClpP1_C without a tag. B, site N mutant.

sequences AY↓NYL and RSL↓YR to AAAAL and RAAAR, respectively. In the resulting strains (ClpP1mC and ClpP1mC'), as well as in the double mutant carrying both mutations (ClpP1mCmC'), the ClpP1_C and ClpP1_{C'} bands were still present (Fig. 5A). The mutation of site C slightly decreased the level of ClpP1_C, with no effect on ClpP1_{C'}, whereas the mutation of site C' slightly increased the level of ClpP1_C and decreased that of ClpP1_{C'} (for quantification, see supplemental Fig. S1). Interestingly, a new immunoreactive band appeared in these mutants (Fig. 5A, asterisk) whose size (39 kDa) corresponds approximately to the C-terminal product of a single cleavage at site N, *i.e.* with neither site C nor C' processing. When both sites were mutated, the intensity of this band increased, whereas that of ClpP1_C and ClpP1_{C'} decreased slightly.

The position of site N cleavage that generates the C terminus of ClpP1_N was established indirectly. Based on the size of ClpP1_N (21 kDa) and on the fact that its migration was unaffected in strain ClpP1HA₁₇₆, cleavage had to occur not too far upstream of position 176. Indeed, a semitryptic peptide (DFSPNQDKDSAN, ending at position 170) was detected in the course of the electrospray ionization-tandem mass spectrometry analysis of the complex, which we thought could mark the C terminus of ClpP1_N. To corroborate this assignment, we mutated this region in Strep-Tag-tagged ClpP1, replacing residues 168–178 with the dipeptide TG. In this strain (ClpP1ΔN-Strep), the apparent molecular mass of ClpP1_N was slightly higher than that in the WT and in the Strep-Tag-tagged strain (Fig. 5B). This suggested that cleavage can still occur, but at a position slightly downstream. This region of the protein therefore appears prone to cleavage, be it at site N or in the vicinity. To attempt to abolish cleavage completely, we introduced the HA tag just downstream of the site N mutation (mutant ClpP1ΔNHA₁₇₆), hoping to change the local conformation of the protein. Indeed, ClpP1_N disappeared (Fig. 3B). However, we

Proteolytic Maturation of Chloroplast ClpP1 Subunit

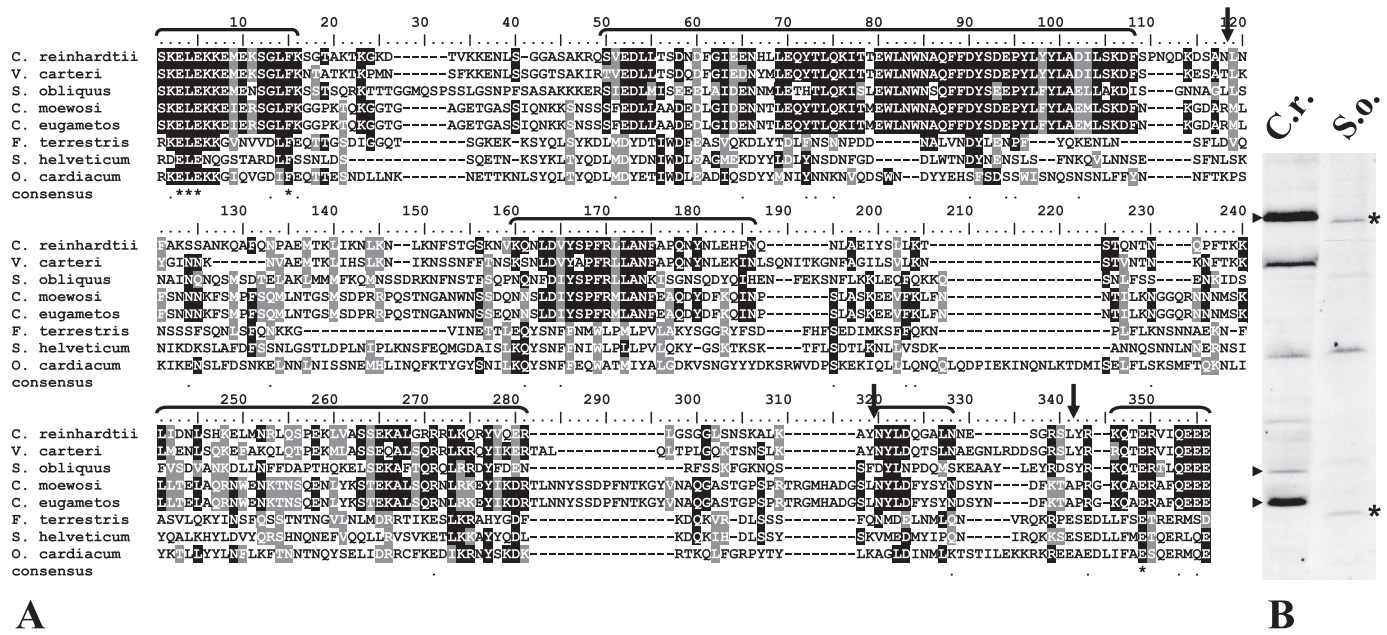


FIGURE 6. IS1 in Chlorophyceae. *A*, shown is an alignment of IS1 sequences from various Chlorophyceae. Cleavage sites are marked by arrows. Six conserved sequence blocks are marked by brackets. *B*, whole cells protein extracts from *S. obliquus* (S.o.) reacted with the anti-ClpP1 antibody. The two bands marked by an asterisk likely correspond to ClpP1_H and ClpP1_N. C.r., *C. reinhardtii*.

observed the appearance of a multiplicity of faint bands of intermediate molecular mass reacting with the anti-HA antibody. Accumulation of ClpP1_H was reduced, as was that of the ClpP complex, as evidenced from the commensurate reduction in ClpR2 amount (data not shown). The mutant protein therefore appeared to be unstable and subject to cleavage at multiple positions. Still, the fact that the mutant could be easily obtained indicates that some functional ClpP complex accumulates in this strain. We concluded that cleavage at or near site N is not strictly necessary for ClpP assembly and function. Other cleavage events can occur when it is prevented, but the protein produced appears less stable. Because of the large reduction in ClpP1 amount in this strain, we were unable to determine whether ClpP1_C and ClpP1_{C'} were still produced in the absence of site N cleavage.

As ClpP1 is itself a peptidase, we speculated that it could be responsible for its own maturation, and we attempted to mutate its catalytic Ser³⁸⁷. We were able to introduce the S387A mutation into the *clpP1* gene, linked to the *aadA* cassette conferring antibiotic resistance. Despite all our efforts, the transformants remained heteroplasmic, *i.e.* retained the WT copy of the *clpP1* gene on a fraction of their plastid chromosomal copies. After up to 12 rounds of subcloning on high concentrations of antibiotics, we achieved homoplasmy for the *aadA* resistance marker in some strains, but only to find that the S387A mutation had been lost, probably because of recombination with the WT *clpP1* copies. This is diagnostic of a lethal mutation (12), and we concluded that not only ClpP1 but also its catalytic activity is essential for cell viability.

IS1 Is Essential in Chlamydomonas—We also investigated whether IS1 is dispensable in *Chlamydomonas*. Using restriction sites present in the plasmid carrying the mutations at both ends of IS1 (S60A/S365A), we removed the intervening sequence, generating a ClpP1 subunit of a canonical length.

Because this changed two residues at the junction between SD1 and SD2 (MEDD↓AKKV), we also generated a plasmid restoring the WT sequence (MEDR↓SKKV), corresponding to a clean excision of IS1. In both cases, we encountered the same problem as with the mutation of the catalytic Ser mentioned above: the mutation could be introduced, by virtue of its linkage to the resistance cassette, but it could not be brought to homoplasmy and was eventually lost by homologous recombination with the WT *clpP1* copy. We concluded that IS1 plays an essential role in the function or biogenesis of the ClpP complex.

IS1 in Chlorophycean Algae—IS1 has been found in the ClpP1 protein not only of *C. reinhardtii* and *C. eugametos* (12) but also of the related multicellular alga *Volvox carteri* (11). These species belong to the family Chlamydomonadales, one of the five families of the class Chlorophyceae. We found an IS1 sequence in all of the Chlorophyceae *clpP1* genes sequenced to date (26–29), including *Chlamydomonas moewusii* (another Chlamydomonadales); *Scenedesmus obliquus* (from the sister group Sphaeropleales); and the more distantly related *Oedogonium cardiacum* (an Oedogoniales), *Floydiella terrestris* (a Chaetophertidales), and *Stigeoclonium helveticum* (a Chaetophorales). In contrast, no IS1 was found in ClpP1 in the other classes of chlorophyte algae: Ulvophyceae (*Pseudodoctonidium akinetum* and *Oltmannsiellopsis viridis*), Trebouxiophyceae (*Leptostira terrestris* and *Chlorella vulgaris*), and Prasinophyceae (*Nephroselmis olivacea*, *Ostreococcus tauri*, and *Ostreococcus lucimarinus*). It was also not found in green algae belonging to the streptophyte lineage and in the early diverging *Mesostigma viride* of uncertain classification. This suggests that IS1 was acquired just before the radiation of the Chlorophyceae. The alignment of these IS1 sequences (Fig. 6A) is consistent with the phylogenetic tree that has been derived by Turmel and co-workers (27) for Chlorophyceae. Another unusual characteristic of Chlorophycean ClpP1 is the presence of a long C-ter-

minal extension compared with higher plants. For example, the sequence of ClpP1 in *F. terrestris* extends 36 amino acids beyond the longest higher plant sequence (*Lycopersicon esculentum*).

Compared with the rest of the ClpP1 protein, sequence variability was high in IS1. Even within the small group formed by Chlamydomonadales and Sphaeropleales, similarity was limited to a few blocks separated by variable regions with many insertions/deletions (Fig. 6A). The sequence around cleavage site C (NYLD) was conserved in Chlamydomonadales and in *S. obliquus*, but that surrounding sites N and C' was not. Still, we found evidence for ClpP1 processing in *S. obliquus*, in which our antibodies recognized two bands upon Western blotting (Fig. 6B, asterisk) that might correspond to ClpP1_H and ClpP1_N. Our antibodies showed no convincing cross-reaction with the other algae tested (*C. moewusii*, *Gloeotilopsis paucicellularis*, *Uronema acuminatum*, *Chlorococum ellipsoïdum*, *Chlorella mirabilis*, *Leptosira obovata*, *Pseudendoclonium basilense*, *Coccomyxa pringsheimii*, and *Coccomyxa rayssiae*) (data not shown).

DISCUSSION

Because ClpP proteins are highly conserved through evolution, the presence in *Chlamydomonas* ClpP1 of an insertion sequence unrelated to any known protein has prompted speculations about its role, fate, and origin (11, 12). It was proposed that IS1 is a new type of protein intron, distinct from inteins. Here, we have ruled out this hypothesis by showing that IS1 is not post-translationally spliced out, but is instead cleaved in at least three places to generate N- and C-terminal fragments that remain associated within the complex. ClpP1_N ends approximately after the first third of IS1, whereas the two alternative C-terminal fragments start near the IS1 end. The N-terminal fragment of ClpP1 (ClpP1_N), referred to previously as ClpP1_L (11, 13), appears responsible for most of the immunoreactivity, explaining why the C-terminal fragments had hitherto escaped our attention. But they can be clearly detected, either when they are marked with a tag, or in a purified ClpP complex, or with an IS1-specific antibody. Note that we consistently observed two C-terminal fragments, which we call ClpP1_C and ClpP1_{C'}. They are produced by cleavage in the C-terminal part of IS1 at either of the two sites C and C'. It is not known whether ClpP1_C and ClpP1_{C'} subunits coexist within one complex or whether the position of the cleavage varies from one complex to the other. The intervening sequence between sites N and C/C' is most likely lost from the complex and rapidly degraded because the anti-HA antibody recognized only ClpP1_H in strain ClpP1HA₁₇₆ (Fig. 3B). When sites C/C' were mutated, a 39-kDa fragment could be detected (Fig. 5A) that probably corresponds to the C-terminal fragment produced by site N cleavage.

These cleavages are more likely to occur on assembled than on unassembled ClpP1 polypeptides. The small fragments may not be able to fold by themselves, especially ClpP1_N, which of the ClpP structural elements retains only two α -helices and one β -strand. Interestingly, only a fraction of the ClpP1 subunits in the complex undergoes processing, as evidenced from the perfect co-migration of ClpP1_H and ClpP1_N in native gels (11). This, together with the large apparent size of the complex (540

kDa, compared with 350 kDa in higher plants), indicates the presence of more than one ClpP1 subunit per complex. Because IS1 processing entails the removal of ~ 15 kDa of protein between N- and C-terminal cleavage sites, the sharpness of the band suggests that the complex comprises a fixed number of unprocessed and processed subunits. Whether a given subunit is processed or not is most likely dictated by the nature of the neighboring subunits. The complex also comprises seven nuclear encoded subunits, some of which are proteolytically active (ClpP4 and ClpP5), whereas others are not (ClpR1–ClpR6).

Can we similarly explain why some ClpP1 subunits are cleaved at site C and others at site C'? A mutation altering the sequence around site C slightly increases the abundance of ClpP1_{C'} while slightly reducing that of ClpP1_C and vice versa. It is tempting to speculate that cleavage occurs at either of these sites, with a probability depending in part on the surrounding amino acid sequence. Because site N cleavage does not seem to be affected in site C/C' mutants (the ClpP1_H/ClpP1_N ratio did not change), we propose that it occurs first. Cleavage at site C or C' may be endoproteolytic or involve processive trimming of the newly formed N terminus. We note that *E. coli* ClpP is able to cleave its own 14-residue propeptide (2) and that the distance between sites C and C' in ClpP1 (15 amino acids) is approximately twice that in the peptides released by ClpP-mediated degradation (6, 30). Thus, we considered the possibility that ClpP performs its own processing. After being processively degraded over a certain length, the polypeptide chain would be stretched between the proteolytic chamber and the already folded SD2 domain, preventing further cleavage. Whether the last cleavage occurs at site C or C' could be influenced by the nature of the residues at this position and by the propensity of the newly generated fragment to remain in the chamber or pop back out. Our efforts to test a possible role of ClpP1 in its own processing proved unsuccessful, as a mutation of the catalytic residue Ser³⁸⁷ remained heteroplasmic. We were surprised to find that not only the ClpP1 protein but also its activity is essential for *Chlamydomonas* survival, despite the presence of other catalytically active subunits in the complex.

Moreover, we found that a deletion of IS1 cannot be brought to homoplasmy, even though the shortened protein is expected to fold normally. IS1 has thus become essential in *Chlamydomonas*, either for the function of the complex or for its biogenesis. In contrast, IS1 cleavage can be severely affected without compromising viability. Mutations at sites C and C' affected the ratio of the two C-terminal fragments and revealed the putative intermediate generated by site N cleavage. Tampering with the site N sequence led either to a shift in the cleavage position (ClpP1 Δ N-Strep) (Fig. 5B) or to a diffuse processing pattern when it was accompanied by the insertion of the HA epitope (ClpP1 Δ NHA₁₇₆) (Fig. 3B). Note that some form of cleavage was always observed, and we cannot rule out that activation of the complex requires one form or another of IS1 removal.

IS1 is predicted to protrude at the apical face of the heptameric ring, where the docking of the chaperone is known to occur. In *Arabidopsis*, small "ClpS" subunits (ClpT is probably a more appropriate nomenclature) may also crowd that surface (31). Assembly of the chloroplast Clp protease appears there-

Proteolytic Maturation of Chloroplast ClpP1 Subunit

fore more complicated than in *E. coli*. This could be related to the diversification of ClpR in photosynthetic organisms. In cyanobacteria, a heptamer made of active ClpP3 has been proposed to interact with another one made of inactive ClpR to form a ClpP3-ClpR complex, distinct from that involving ClpP1 and ClpP2 (32). Cyanobacterial ClpP3 has given rise to chloroplast ClpP1 and ClpR2, whereas ClpR has diversified to produce ClpR1, ClpR3, and ClpR4. In Chlorophyceae, ClpR6 is derived from ClpP6 (11). However, some reorganization of the subunits must have occurred along the way, as in *Arabidopsis*, one heptameric ring containing ClpP1 and ClpR1-ClpR4 appears to be associated with another one made of ClpP3-ClpP6 (33). Whatever the organization of the *Chlamydomonas* complex, it is reasonable to assume that the presence of IS1 on at least one of the apical surfaces will impose strong constraints on the interaction with the chaperone and that IS1 processing represents a manner of activation of the ClpP complex.

The above discussion underscores how little we understand of the function and origin of IS1. Like many biological processes (mRNA editing, inteins, etc.), it seems to provide no competitive advantage, yet it has been conserved through the evolution of Chlorophyceae and has even become essential in *Chlamydomonas*. We must therefore take up a different point of view and ask how IS1 came about. More precisely, because ClpP1 is essential in all photosynthetic organisms (12, 34, 35), how did the system survive the introduction of a large additional domain disrupting a crucial interaction? And what changes occurred in the Clp system that now make IS1 essential?

First, we note that the interaction with the chaperone can be maintained if all ClpP1 subunits are on the same side of the complex (a hypothesis supported by the data in *Synechocystis* and *Arabidopsis*). The peptidase would simply have an active and an inactive surface with respect to chaperone docking and substrate entry. In this view, the function of IS1 processing is not to allow docking of the chaperone. The role of IS1 and its processing may instead be in regulating assembly of the complex.

The IS1 sequence shows no similarity to other proteins in the data bases. This suggests that it results from the insertion within *clpP1* of a DNA fragment that had no specific coding function or lost it in the process. Apparently, *clpP1* is not the sole chloroplast gene of green algae that had to cope with such an adventitious insertion. Chloroplast genome sequences from a variety of Chlorophyta (26-29, 36-38) show that, whereas the most ancestral group (Prasinophyceae, exemplified by *Mesostigma*) has retained a streptophyte-like organization, the more evolved Trebouxiophyceae, Ulvophyceae, and Chlorophyceae (the "UTC" group) are highly rearranged, with general scrambling of gene order and many gene losses. This correlates with a steady increase in the number and extent of short dispersed repeats (39). Simultaneously, a general increase in ORF length can be observed (35), probably caused by the introduction of "junk" DNA within permissive regions of the genes. Examples of such insertion sequences are presented in supplemental Figs. S2-S5 for CemA, RpoB, RpoC1, and RpoC2. Some of these insertions appear long enough to fold as an independent domain within the host protein, just as IS1 does in ClpP1. However, except for the maintenance of ORF continuity, most

show a low degree of sequence conservation, even compared with ClpP1 IS1. This suggests that they have not acquired specific functions. We speculate that many will turn out to be cleaved post-translationally.

Insertion sequences can also evolve into gene splitting. For example, RpoB2 contains an insertion sequence in Ulvophyceae and in *Chlorella*, which independently gave rise to gene splitting in Chlorophyceae and in *Leptosira*. The gene-splitting sites in *C. reinhardtii* RpoC1 and *C. moewusii* RpoC2 also coincide with the sites of large insertion sequences in other Chlorophyta.

In conclusion, the IS1 sequence found in the ClpP1 protein of *C. reinhardtii* and other Chlorophyceae is not a protein intron. It probably derives from a DNA insertion into the *clpP1* gene of an ancestral Chlorophyceae. In a fraction of the ClpP1 subunits, it is proteolytically processed, probably after the assembly of the complex. During the course of its evolution, IS1 has acquired new functions that today make it an essential domain of an essential protein.

Acknowledgments—We thank Lisa Granelli and Sophie Fleurdépine for help with the initial constructions, J.-D. Rochaix (University of Geneva) for the pHAtag plasmid, and Pierre Cardol for other algal strains. We thank J. d'Alayer for determination of N-terminal sequences, Y. Choquet for critical reading of the manuscript, and all members of UMR7141 for stimulating discussions.

REFERENCES

- Katayama, Y., Gottesman, S., Pumphrey, J., Rudikoff, S., Clark, W. P., and Maurizi, M. R. (1988) *J. Biol. Chem.* **263**, 15226-15236
- Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., and Gottesman, S. (1990) *J. Biol. Chem.* **265**, 12536-12545
- Gottesman, S., Clark, W. P., de Crecy-Lagard, V., and Maurizi, M. R. (1993) *J. Biol. Chem.* **268**, 22618-22626
- Ortega, J., Lee, H. S., Maurizi, M. R., and Steven, A. C. (2004) *J. Struct. Biol.* **146**, 217-226
- Wang, J., Hartling, J. A., and Flanagan, J. M. (1997) *Cell* **91**, 447-456
- Thompson, M. W., Singh, S. K., and Maurizi, M. R. (1994) *J. Biol. Chem.* **269**, 18209-18215
- Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. (1998) *Genes Dev.* **12**, 1338-1347
- Hoskins, J. R., Kim, S. Y., and Wickner, S. (2000) *J. Biol. Chem.* **275**, 35361-35367
- Singh, S. K., Grimaud, R., Hoskins, J. R., Wickner, S., and Maurizi, M. R. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8898-8903
- Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T., and Baker, T. A. (2000) *Mol. Cell* **5**, 639-648
- Majeran, W., Friso, G., van Wijk, K. J., and Vallon, O. (2005) *FEBS J.* **272**, 5558-5571
- Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M., and Liu, X. Q. (1994) *Mol. Gen. Genet.* **244**, 151-159
- Majeran, W., Wollman, F. A., and Vallon, O. (2000) *Plant Cell* **12**, 137-150
- Perler, F. B. (2002) *Nucleic Acids Res.* **30**, 383-384
- Wang, S., and Liu, X. Q. (1997) *J. Biol. Chem.* **272**, 11869-11873
- Harris, E. H. (1989) *The Chlamydomonas Source Book: A Comprehensive Guide to Biology and Laboratory Use*, Academic Press, San Diego, CA
- Goldschmidt-Clermont, M. (1991) *Nucleic Acids Res.* **19**, 4083-4089
- Higuchi, R. (1990) in *PCR Protocols: A Guide to Methods and Application* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., eds) pp. 177-183, Academic Press, New York
- Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., Johnson, A. M., Jones, A. R., Randolph-Anderson, B. L., Robertson, D., Klein, T. M., and

Proteolytic Maturation of Chloroplast ClpP1 Subunit

- Shark, K. B. (1988) *Science* **240**, 1534–1538
20. Kuras, R., and Wollman, F. A. (1994) *EMBO J.* **13**, 1019–1027
21. Werner, R., and Mergenhagen, D. (1998) *Plant Mol. Biol. Rep.* **16**, 295–299
22. Piccioni, R. G., Bennoun, P., and Chua, N. H. (1981) *Eur. J. Biochem.* **117**, 93–102
23. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354
24. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001) *Methods* **24**, 218–229
25. Perler, F. B. (2005) *IUBMB Life* **57**, 469–476
26. Turmel, M., Brouard, J. S., Gagnon, C., Otis, C., and Lemieux, C. (2008) *J. Phycol.* **44**, 739–750
27. Bélanger, A. S., Brouard, J. S., Charlebois, P., Otis, C., Lemieux, C., and Turmel, M. (2006) *Mol. Genet. Genomics* **276**, 464–477
28. Brouard, J. S., Otis, C., Lemieux, C., and Turmel, M. (2008) *BMC Genomics* **9**, 290
29. de Cambiaire, J. C., Otis, C., Lemieux, C., and Turmel, M. (2006) *BMC Evol. Biol.* **6**, 37
30. Thompson, M. W., and Maurizi, M. R. (1994) *J. Biol. Chem.* **269**, 18201–18208
31. Peltier, J. B., Ripoll, D. R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J., Giacomelli, L., Pillardy, J., and van Wijk, K. J. (2004) *J. Biol. Chem.* **279**, 4768–4781
32. Stanne, T. M., Pojidaeva, E., Andersson, F. I., and Clarke, A. K. (2007) *J. Biol. Chem.* **282**, 14394–14402
33. Sjögren, L. L., Stanne, T. M., Zheng, B., Sutinen, S., and Clarke, A. K. (2006) *Plant Cell* **18**, 2635–2649
34. Kuroda, H., and Maliga, P. (2003) *Nature* **425**, 86–89
35. Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T., and Hashimoto, T. (2001) *Plant Cell Physiol.* **42**, 264–273
36. de Cambiaire, J. C., Otis, C., Turmel, M., and Lemieux, C. (2007) *BMC Genomics* **8**, 213
37. Pombert, J. F., Lemieux, C., and Turmel, M. (2006) *BMC Biol.* **4**, 3
38. Lemieux, C., Otis, C., and Turmel, M. (2000) *Nature* **403**, 649–652
39. Lemieux, C., Otis, C., and Turmel, M. (2007) *BMC Biol.* **5**, 2