

Identification of a 350-kDa ClpP Protease Complex with 10 Different Clp Isoforms in Chloroplasts of *Arabidopsis thaliana**

Received for publication, November 20, 2000, and in revised form, January 17, 2001
Published, JBC Papers in Press, January 26, 2001, DOI 10.1074/jbc.M010503200

Jean-Benoit Peltier^{‡§}, Jimmy Ytterberg^{‡§¶}, David A. Liberles^{‡¶}, Peter Roepstorff^{**‡‡},
and Klaas Jan van Wijk^{‡ §§}

From the [‡]Department of Biochemistry, Arrhenius Laboratories, Stockholm University, S-10691 Stockholm, Sweden, the [¶]Stockholm Bioinformatics Center, Stockholm University, S-10691 Stockholm, Sweden, and the ^{**}Department of Molecular Biology, Odense University, DK-5230 Odense M, Denmark

A 350-kDa ClpP protease complex with 10 different subunits was identified in chloroplast of *Arabidopsis thaliana*, using Blue-Native gel electrophoresis, followed by matrix-assisted laser desorption/ionization time-of-flight and nano-electrospray tandem mass spectrometry. The complex was copurified with the thylakoid membranes, and all identified Clp subunits show chloroplast targeting signals, supporting that this complex is indeed localized in the chloroplast. The complex contains chloroplast-encoded pClpP and six nuclear-encoded proteins nCpP1–6, as well as two unassigned Clp homologues (nCpP7, nCpP8). An additional Clp protein was identified in this complex; it does not belong to any of the known Clp genes families and is here assigned ClpS1. Expression and accumulation of several of these Clp proteins have never been shown earlier. Sequence and phylogenetic tree analysis suggests that nCpP5, nCpP2, and nCpP8 are not catalytically active and form a new group of Clp higher plant proteins, orthologous to the cyanobacterial ClpR protein, and are renamed ClpR1, -2, and -3, respectively. We speculate that ClpR1, -2, and -3 are part of the heptameric rings, whereas ClpS1 is a regulatory subunit positioned at the axial opening of the ClpP/R core. Several truncations and errors in intron and exon prediction of the annotated Clp genes were corrected using mass spectrometry data and by matching genomic sequences with cDNA sequences. This strategy will be widely applicable for the much needed verification of protein prediction from genomic sequence. The extreme complexity of the chloroplast Clp complex is discussed.

The chloroplast has been predicted to contain approximately 2000–2500 proteins (1, 2), and many of these proteins are

* This work was supported in part by the Swedish Agricultural Research Council, the Swedish Foundation Strategic Research, a grant for two-dimensional equipment from the Swedish National Research Council, and a grant for mass spectrometers from the Hasselblad Foundation (to K. J. v. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Present address: Dept. of Plant Biology, Cornell University, Ithaca, NY 14853.

¶ Supported by the program Cell Factory for Functional Genomics of the Swedish Foundation for Strategic Research.

‡‡ Member of the Center for Experimental Bioinformatics sponsored by the Danish National Research Foundation.

§§ To whom correspondence should be addressed. Present address: Dept. of Plant Biology, Emerson Bldg., 3rd Fl., Tower Rd., Cornell University, Ithaca, NY 14853. Tel.: 607-254-1211; Fax: 607-255-5407; E-mail: kv35@cornell.edu.

likely to function as stable or transient protein complexes (3). In this study we report on the identification of a 350-kDa ClpP protease complex with 10 different subunits that was released from the thylakoid membranes of *Arabidopsis thaliana* by Yeda press treatment. The complex was identified through Blue-Native gel electrophoresis (4), followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)¹ and nano-electrospray tandem mass spectrometry (ESI-MS/MS) (for reviews see, e.g., Refs. 5 and 6).

Clp proteins (for caseinolytic protease) were first identified in *Escherichia coli* (7, 8). In *E. coli*, a Clp complex with a mass of ~300 kDa was shown to be an ATP-dependent serine type protease consisting of a catalytic core composed of two apposite heptameric rings of ClpP protease subunits. All 14 ClpP subunits are identical and are encoded by one gene. This catalytic core can be flanked at one or both ends by a hexamer of regulatory ClpA/X chaperone subunits, belonging to the family of ATP-dependent HSP100 proteins (9). This family, is divided into class I with five subclasses (A, B, C, D, and E) and II with four subclasses (M, N, X, and Y). Proteins in class I contain two nucleotide-binding domains (NBDs), and proteins in the second class are shorter in length, containing only one NBD and a conserved C-terminal region (10).

The structure of the heptameric ClpP ring in *E. coli* has been determined at 2.3-Å resolution and resembles a hollow, solid-walled cylinder ~51 Å in diameter with only two axial openings (11). The narrow width of the pore opening (~10 Å) suggests that protein substrates enter the proteolytic chamber in a largely unfolded conformation (11). The ClpP complex alone degrades only peptides of fewer than 6 amino acid residues, but when the hexameric chaperone rings are attached to ClpP, the complex can degrade longer polypeptides and proteins (12). The proteolytic core can also be formed by ClpQ (HslV) proteins, which are otherwise unrelated to ClpP, both in terms of amino acid sequence and mode of proteolytic action (13).

The Clp family in higher plants (*A. thaliana*) consists of at least the plastid-encoded pClpP gene, six homologous nuclear-encoded nClpP genes, and the nuclear encoded chaperones ClpB, -C, -X, and ClpD, also named ERD1 (for early responsive to dehydration). Information concerning chloroplast Clp proteins has been reviewed in Refs. 14 and 15. No ClpQ homologue has been found in any plant species. Cyanobacteria have an additional ClpP homologue, named ClpR, but, as this homo-

¹ The abbreviations used are: MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MS, mass spectrometry; ESI, electrospray; Rubisco, ribulose-bisphosphate carboxylase/oxygenase; PAGE, polyacrylamide gel electrophoresis; NBD, nucleotide-binding domain; EST, expressed sequence tag; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

logue does not have an active catalytic triad, it is believed to be a regulatory subunit (45).

There is evidence for chloroplast localization of pClp, nClpP1, ClpC, and ERD1, from either *in vitro* import into isolated chloroplasts or Western blotting and immunodetection (16–19). pClpP and ClpC were reported to be constitutively expressed in both photosynthetic and nonphotosynthetic tissue in equimolar amounts. Their protein levels also remained unchanged during various stress treatments, such as heat shock and dehydration (16). Accumulation of transcripts of ERD1, nClp3, and nClp5 was shown to be up-regulated during drought stress and other stresses, during long term dark treatment, as well as during senescence, suggesting that these gene products function during natural senescence (17, 20). During senescence, the expression of pClpP, nClpP1, nClpP2, nClpP4, and nClpP6, on the other hand, seems to stay unchanged (17). Coimmunoprecipitation showed that pClpP and nClpP6 were associated with ClpC (21, 22), whereas immunoblotting of size-fractionated chloroplast stroma indicated that ClpC, pClpP and nClpP6 migrated in a range of different sized complexes (between 160 and 1700 kDa) (22). It was thus postulated that heterocomplexes such as ERD1/ClpP1 or ClpC/ClpP1, or ERD1/CplC/pClpP (etc.) could exist (17, 22).

The physiological role for the Clp proteins in chloroplasts is to degrade misfolded or unassembled proteins in an ATP-dependent manner (23–25). In the green algae *Chlamydomonas*, the pClpP gene was shown to be essential (26), whereas reduced expression levels of pClpP restricted the ability of the cells to adapt to high CO₂ levels and also suggested that ClpP is involved in proteolytic disposal of fully or partially assembled cytochrome *b₆f* (25). Inactivation of ClpP in *E. coli* and *Synechocystis* is not lethal (27, 28).

Despite these studies, no chloroplast Clp complex has been isolated and its subunits identified and characterized. Such isolation and identification of the subunits is needed to determine the stoichiometry and expression of the different isoforms and to understand the role of the different gene products and the relationship to stress and chloroplast development and senescence.

In this study, we identified a chloroplast Clp complex of 350 kDa using nondenaturing Blue-Native gel electrophoresis (4). Thylakoid membrane-associated complexes were separated by Blue-Native PAGE, and each complex was then separated into their individual subunits by SDS-PAGE and visualized by Coomassie and silver staining. The Clp complex separated into seven strong and two weak protein spots, and all protein spots were successfully analyzed by MALDI-TOF MS and ESI-MS/MS. Analysis of the mass spectrometry data showed that the isolated Clp complex contained 10 different Clp proteins, including pClpP, 5 nuclear-encoded ClpP proteins, 3 nuclear-encoded ClpR proteins, as well as 1 novel Clp protein, ClpS1.

The nomenclature of the Clp proteins in higher plants is confusing, with several proteins going under different names. In agreement with a suggestion for nomenclature of chloroplast proteases (45), we have (re)named a number of the Clp genes, based on presence or absence of functional domains and confirm the homology to Clp protein in *E. coli* and *Synechocystis* PCC6803.

EXPERIMENTAL PROCEDURES

Plant Growth—*A. thaliana*, ecotype Columbia, was grown on soil in a temperature-controlled growth chamber under a 10/14-h light dark cycle, at 21 °C/16 °C light/dark temperatures at a light intensity of about 100 μmol of photons m⁻² s⁻¹.

Isolation of Chloroplast Clp Complexes by Two-dimensional Gel Electrophoresis—Leaves from *Arabidopsis* were homogenized in ice-cold medium A (50 mM Hepes-KOH, pH 7.5, 100 mM sorbitol, 5 mM MgCl₂, 50 μg/ml Pefabloc, 2 μg/ml antipain, 2 μg/ml leupeptin), filtered through four layers of Miracloth (22 μm), and centrifuged for 3 min at 9000 × *g*. Thylakoid pellets were resuspended in medium A and cen-

trifuged on a Percoll gradient (0–90%) in medium A for 4 min at 3000 × *g*. The thylakoid band was collected and washed three times in the same medium A. Thylakoids were passed twice through the Yeda press (at 100 bar), and thylakoid membranes were subsequently removed by ultracentrifugation (60 min at 150,000 × *g*). The membrane-free supernatant was collected and concentrated to ~20 mg protein·ml⁻¹. Samples were then mixed with an equal volume of 50 mM caproic acid, 50 mM Bis-Tris, pH 7.0, 15% glycerol, 0.2% Coomassie Brilliant Blue G-250, and separated on gradient (4–12% acrylamide) Blue-Native gels (4) at 4 °C. Molecular mass complexes ranging from 67 to 669 kDa (Amersham Pharmacia Biotech) were loaded for calibration of the gels. Gel lanes were then cut out and equilibrated twice for 20 min in SDS solubilization buffer (containing 6 M urea, 50 mM Tris-HCl, pH 6.8, 30% glycerol and 2% SDS; 5 mM tributylphosphine). Proteins were separated on 8–16% acrylamide Tricine gels (29). Gels were stained using Coomassie Brilliant Blue R-250 or silver nitrate (30, 31).

Mass Spectrometry, Data Base Searching, Sequence Analysis, and Localization Prediction—Stained protein spots were excised from the gel, washed, reduced, alkylated using iodoacetamide, and digested with modified trypsin (Promega), according to Ref. 31. The peptides were then extracted and dissolved in 20 μl of 5% formic acid and applied to the MALDI-TOF target plate by the dried droplet method using α-cyano-4-hydroxycinnamic acid as matrix. When necessary, the samples were concentrated using microcolumns (32) and eluted directly onto the MALDI target. The mass spectra were obtained using a MALDI-TOF mass spectrometer (REFLEX II from Bruker Daltonics and Voyager-DE-STR from PerSeptive Biosystems Inc.). The spectra were annotated with the program “m/z” from Proteometrics and internally calibrated using tryptic peptides from autodigestion. The latest versions of the NCBI nonredundant data base were searched with the resulting peptide mass lists, using the search engine ProFound. The search strategy was in principle as described previously (33).

To further analyze the samples, the remainder of the extracted peptides were desalted and concentrated on microcolumns (Poros R2, PE Biosystems) and eluted directly into nanoelectrospray needles (Protana A/S, Odense, Denmark) with 1.2 μl of 50% MeOH and 1% formic acid (34). The spectra were acquired on an electrospray tandem mass spectrometer (Q-TOF Micromass). The instrument was calibrated with 1 μg/μl NaI in 50% isopropanol. Alternatively the spectra were used to search the public data bases with the program Mascot. Alternatively, the MS/MS spectra were interpreted using MassLynx and PepSeq (Micromass) and were used to search different public data bases using FASTA3.

Predictions for chloroplast localization, chloroplast, and luminal transit peptides were made using the search engines TargetP and Predotar.

Construction of Phylogenetic Trees and Multi-alignments—First, the sequences were grouped into clades of those where homologous relationships could be established, using the union find algorithm applied with 20% and 30% identity threshold. This type of approach has been employed in establishing clusters of genes or proteins in various data bases (35, 36).

For the tree building of all the *Arabidopsis* Clp genes, two distinct clades were identified. In both clades, sequences were identified using a 20% cutoff, but not a 30% cutoff and were excluded from further phylogenetic study. For both clades, multiple sequence alignments were constructed using ClustalW. Additional multi-alignments were done using MultAlin. Using these multiple sequence alignments in PHYLIP, phylogenetic trees were calculated using maximum likelihood (Molphy) parsimony and two distance based methods (neighbor joining and UPGMA) with bootstrap values to test the statistical robustness of the trees (37). Relationships that were present in all of the tree methodologies supported by bootstrap values >95% by at least one methodology are depicted, whereas more ambiguous relationships are presented as start clusters. A box is placed over the central start, where homology cannot be established based upon sequence alone. A second tree using ClpP/R sequences from *Arabidopsis*, *Synechocystis*, and *Synechococcus* was built using ClustalW alignment and both neighbor-joining and maximum likelihood trees, which were alike.

RESULTS

Protein Purification—To identify protein complexes that are peripherally attached to thylakoid membranes, we first purified fairly large quantities of thylakoid membranes (membranes containing about 100 mg of chlorophyll, equivalent to ~800–900 mg of protein) from mature leaves of *Arabidopsis thaliana*. The chloroplast stroma is reported to contain 15–30 mM Mg²⁺, which is needed to keep the thylakoid membranes in

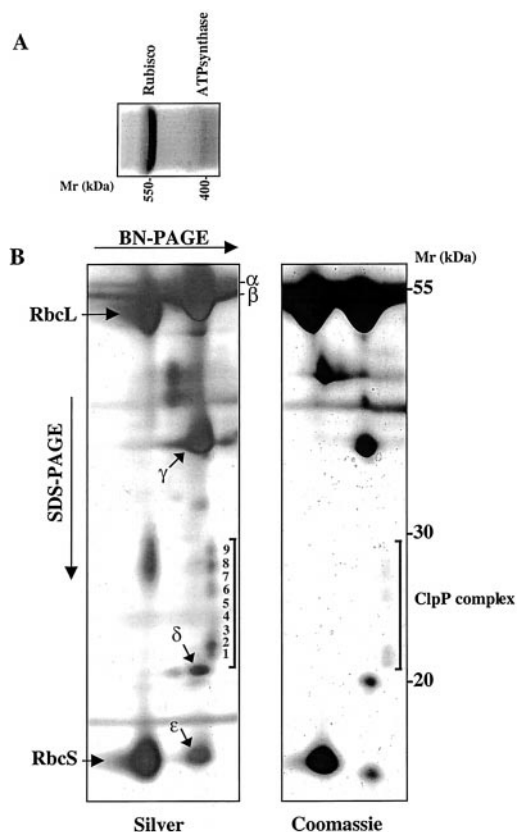


FIG. 1. Separation of chloroplast-localized protein complexes Rubisco, peripheral ATP synthase (CF1), and a Clp complex on Blue-Native PAGE, followed by SDS-PAGE. The complexes were copurified with *Arabidopsis thaliana* thylakoid membranes and released by Yeda press. **A**, the region of 350–600 kDa of the Blue-Native PAGE gel shows the presence of Rubisco and the peripheral ATP synthase complex. **B**, the silver- and Coomassie-stained second dimension SDS gel of the 350–600-kDa region showing the two subunits of the 550-kDa Rubisco complex (RbcL and RbcS), five subunits of the 400-kDa peripheral ATP synthase complex (CF1 α , - β , - γ , - δ , and - ϵ), and nine spots for a complex (the Clp complex), at 350 kDa.

its physiologically stacked state and to stabilize several known complexes, such as ribosomes. Therefore, the thylakoid membranes were kept in a medium containing $MgCl_2$ all throughout the preparation, as we expected that this could enhance binding and stability of membrane-associated protein complexes. The thylakoid membranes were purified on linear Percoll gradients to minimize contamination with other membranes or organelles.

To release peripherally attached thylakoid complexes, we attempted treatment with very low levels of nonionic detergents and treatment by Yeda press, both followed by differential centrifugation. We will not further report on the detergent extraction since this did not result in the purification of significant amounts of novel stable peripheral complexes. Instead, we report on the purification of complexes using Yeda press.

Thylakoid membranes were fragmented by Yeda press treatment, and the membranes were subsequently removed by ultracentrifugation. The remaining supernatant was concentrated and loaded on Blue-Native gels to separate the native protein complexes according to their molecular mass. High molecular mass markers were run in parallel to estimate the molecular masses of the peripheral complexes (Fig. 1A). At the end of the electrophoretic run, three focused protein complexes were visible in the region of 350–600 kDa and their estimated molecular masses were 400 and 550 kDa (Fig. 1A). Gel lanes were then excised and incubated in SDS, and the complexes were separated into their individual subunits by SDS-PAGE. Fig. 1 (B and C) show a silver-stained and Coomassie-stained gel of the 350- and

600-kDa region. In this molecular mass range, we detected two dominant complexes and one complex of lower abundance. Protein spots of each of the three complexes were picked from the gels and prepared for mass spectrometry analysis.

Mass Spectrometry and Protein Identification—Protein spots from the gels were digested with trypsin and the generated peptides were extracted and analyzed by MALDI-TOF MS. To identify the protein(s) in each spot, the experimentally determined peptide masses, were used to search the public sequence data bases (see “Experimental Procedures”). The largest complex was Rubisco (550 kDa) with the large (RbcL) and small subunit (RbcS), and the complex at 400 kDa was the ATP synthase complex CF1 (data not shown) (Fig. 1).

In the third, low abundant complex of ~350 kDa, we could distinguish clearly seven protein spots after silver staining. These seven spots, as well as the gel material in between the spots were excised, such that nine slices, covering the range of 20–30 kDa, were collected. MALDI-TOF MS spectra were obtained from all of them, resulting in the identification of at least one protein per gel slice (Table I). From the MALDI-TOF MS data, it became clear that the 350-kDa complex contains several Clp proteins, suggesting we had purified a chloroplast Clp complex with several isoforms. To obtain further confirmation of these identities and to search for additional proteins in the spots, the remainder of each peptide extract was analyzed by nano-electrospray MS/MS to generate sequence tags. The precursor and fragment ions masses were then used to search NCBI directly (see “Experimental Procedures”). Additionally, all MS/MS spectra were interpreted into amino acid sequences, to search the *Arabidopsis* data base with FASTA and thus to allow for errors in gene annotation (see below and “Discussion”).

To illustrate this identification process, the MALDI-TOF MS spectrum and some of the MS/MS spectra obtained from spot 2 is shown in Fig. 2. Fig. 2A shows the MALDI-TOF spectrum from spot number 2, which contained a mixture of four different Clp proteins (Table I). In the spectrum we annotated 70 monoisotopic masses in the m/z region from 800 to 3300. Nine experimental peptide masses matched to the ClpS1 protein (formerly annotated as hypothetical protein) and covered in total 44% of the primary sequence. The more intense matching peptides are indicated in the spectrum (Fig. 2A). Ten of the peptides in the MALDI spectrum matched to the nClpP5 protein, providing a sequence coverage of 34%. Also ClpP1 (formerly pClpP) protein was identified with confidence in this protein spot, as five peptides matched with 34% coverage. Finally, the MALDI data indicated the possible presence of a fourth Clp protein, ClpP6, but since only three experimental peptides matched (with 17% coverage), this identification clearly needed confirmation (see below). None of the experimental peptide masses matched to the predicted signal peptides of ClpS1, nClpP1 and nClpP6, in agreement with the notion that processing of the chloroplast transit peptide of all nuclear encoded proteins occurs rapidly (38).

After the MALDI-TOF analysis the remainder of the extracted peptides were analyzed by tandem MS. Fig. 2B shows the full scan spectrum in the MS mode (m/z range of 400–1200) of spot 2. In the electrospray spectrum doubly charged, as well as triply charged peptides, were detected (Fig. 2B). Several of these multiple charged peptides were further fragmented along the protein backbone by collision-induced dissociation. Fig. 2C shows the MS/MS spectrum of the doubly charged peptides at m/z 848.4 $[M + 2H]^{2+}$. The complete y ion series (from y1 to y15) could be assigned, as indicated (y ions contain the C terminus of the precursor ion; for a detailed explanation, see, e.g., Refs. 5 and 39). The experimentally determined peptide sequence (by reading the sequence from y15 to y1) matched

TABLE I
 Identification and analysis of the Clp proteins in the 350-kDa complex by mass spectrometry
 Proteins were identified by a combination of MALDI-TOF MS and ESI-MS/MS. The spot numbers refer to the two-dimensional gel in Fig. 1B. The localization and stromal cleavage site of the identified proteins was predicted by the search engine TargetP.

Spot no. ^a	Identity (previous names)	Size in ad ^b (chromosome)	Accession no. ^c	MALDI-TOF		Sequences ^f	MS/MS	Precursor ion ^g	Localization and cleavage site prediction	
				Cover ^d 50 ppm	No matching ^e peptides				P ^h	CS ⁱ
1 (22)	ClpP1 (pClpP)	196 (plastid)	7525058 ^j	33	6	VMIHQPASSFYEAQTGEFII EAEEELK		1032.9 (3)	Plastid-encoded	
	ClpS1 (nClpC like)	241 (+14) (IV)	7487579 ^k	44	9	SDMYFFSPEHPPLTEPAQK EYGLIDGVLNPLK		746.3 (3) 789.4 (2)	Plastid: 0.974	63
	ClpP5 (nClpP1)	298 (I)	4887543 ^l	34	10	ANLDGYLAHTGOSLEK FQSTII SOLFOYR		627.3 (3) 765.4 (2)	Plastid: 0.901	62
2 (22)	ClpP1 (pClpP)	196 (plastid)	7525058 ^j	34	5	VMIHQPASSFYEAQTGEFII EAEEELK		1032.9 (3)	Plastid-encoded	
	ClpP6 (nClpP6)	271 (I)	2827888 ^m	18	3	SPGEGDTSWVDIYNR IIFIGQPIAQAQR		848.4 (2) 834.5 (2)	Plastid: 0.930	62
	ClpP5 (nClpP1)	298 (I)	4887543 ^l	26	4	INQDTRFFMSAK FQSII SQLFOYR ANLNGYLAYHGQSLEK		568.6 (3) 765.4 (2) 940.5 (2)	Plastid: 0.901	62
3+4 (23-24)	ClpP2 (nClpP7)	241 (+34) (V)	9759372			EYGLIDGVLNPLK VMDALNELYVK AFGLIIDEVIDERPLELVK SLENAVMIHQPSGGYSGQAK		789.4 (2) 675.4 (2) 686.0 (3) 720.4 (3)	Mito: 0.593 Plastid: 0.268	32
5 (25)	ClpP2 (nClpP7)	241 (+34) (V)	9759372	39	6	AFGLIIDEVIDERPLELVK WVDALNELYVK AFGLIIDEVIDER		686.4 (3) 675.4 (2) 560.8 (2)	Mito: 0.593 Plastid: 0.268	32
6 (26)	ClpR2 (nClpP2)	279 (I)	5360589 ⁿ	30	8	GLAYLAGFLLAAGFK ^z EEGTWQWVDIWNALYR VIFIGQNI DEEFNQI L ATMLYLD TFLD DSSRR		820.5 (2) 779.4 (3) 1216.0 (3)	Plastid: 0.724	54
7 (27)	ClpP4 (nClpP4)	299 (V)	5360593 ^o	56	8	GAESDVMGLLLR IMLHQP LGGASQAIDVEIQAK NNVTSII IAGCTS VNYEEI SKDPMK VPS SGLMPASDVLIR		638.8 (2) 764.7 (3) 703.9 (2) 490.3 (3) 779.4 (2)	Plastid: 0.963	60
8 (28)	ClpP4 (nClpP4)	299 (V)	5360593 ^o	32	9	GAESDVMGLLLR		638.9 (2)	Plastid: 0.963	60
	ClpR3 (nClpP8)	331 (+39) (I)	2342674	22	9	VPS SGLMPASDVLIR		779.5 (2)	Plastid: 0.863	43
	ClpR1 (nClpP5)	387 (I)	5360595 ^p	24	9				Plastid: 0.933	41
9 (29)	ClpR3 (nClpP8)	331 (+39) (I)	2342674	19	7	DDGETVGMSESEGFAYDSLMLK		940.3 (3)	Plastid: 0.863	43
	ClpR1 (nClpP5)	387 (I)	5360595 ^p	24	9	TAPPDLPSILLDAR ELDANTEYYIELLAK		740.0 (2) 893.1 (2)	Plastid: 0.933	41
	ClpP3 (nClpP3)	331 (I)	5360591 ^q	34	7	DNFLNPEAKEYGLI DAVIDDGK		856.5 (3)	Plastid: 0.904	71

^a Experimental molecular mass.
^b Predicted length of the protein in number of amino acids; in parentheses is correction based on comparison to EST sequences and MS data (see Figs. 3 and 5). The chromosome localization is indicated in Roman numbers.
^c Accession number, but there are also other accession numbers for (nearly) identical proteins.
^d Percentage of coverage of the protein sequence by the matching peptides.
^e Number of matching peptides at mass accuracy of 50 ppm, allowing maximally 1 missed cleavage per peptide.
^f Peptide sequence of the precursor ion.
^g m/z value of the precursor ion, and the charge (z) is indicated in parentheses.
^h Chance (P) that the predicted localization is correct.
ⁱ Amino acid position of the predicted cleavage site (CS).
^j Also 6685310 and 5360576.
^k Also 7268455 and 2245061.
^l Also 5360579 but is 18 aa shorter.
^m Also 4835782.
ⁿ Also 8778627.
^o Also 9758729 (but with 9-amino acid insert) and 4454056 (64 amino acids missing at N terminus).
^p Also 8569101.
^q Also 3599968 (but 5-amino acid difference in N terminus).
^r The protein sequence does not fit entirely with the experimental mass of the precursor ion, most likely due to an error in the predicted protein sequence or because of a post-translational modification.

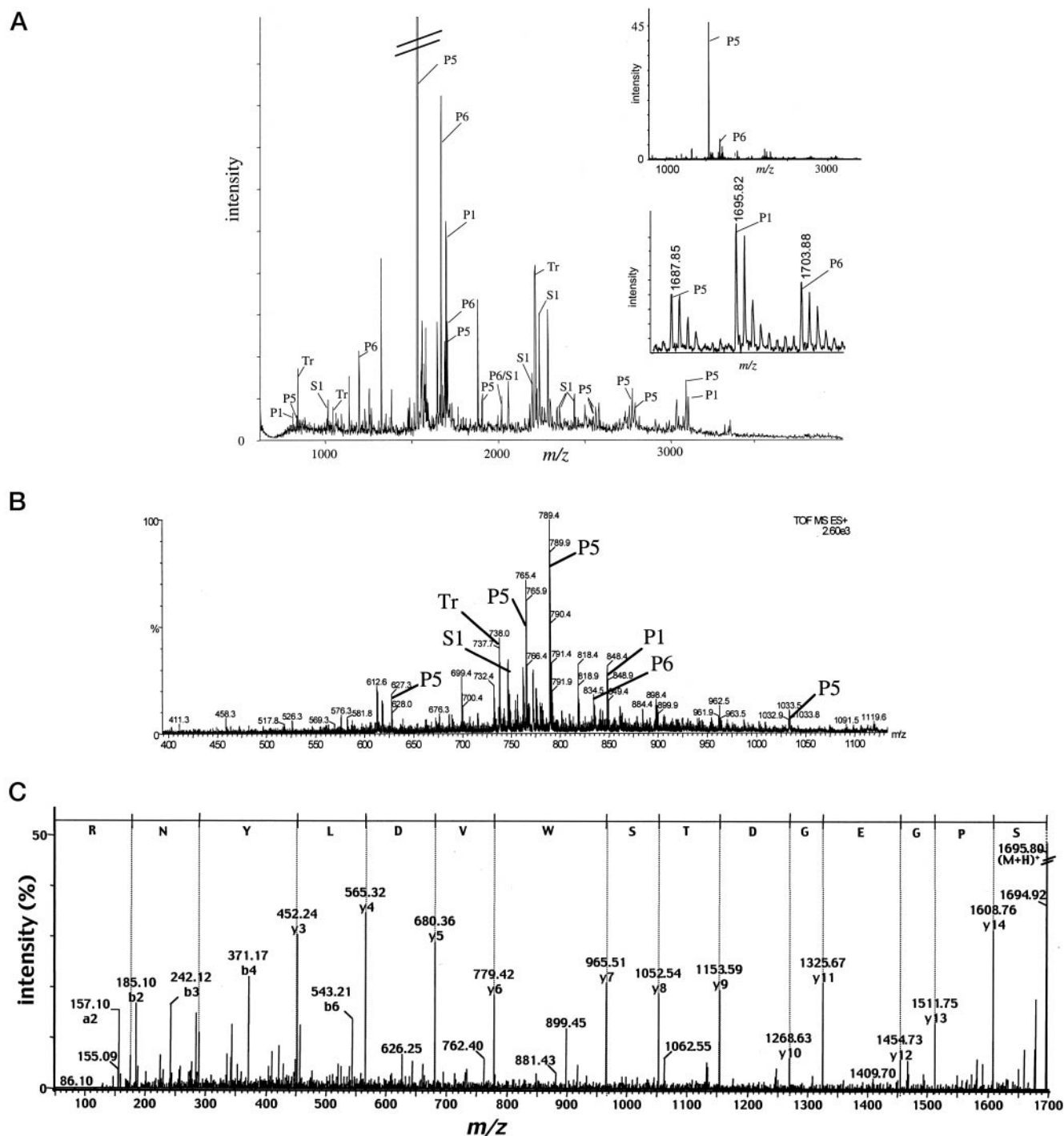


FIG. 2. Mass spectrometry analysis of spot number 2 of the Clp complex. A, MALDI-TOF MS spectrum of the peptides generated by tryptic digestion of the protein spot 2. The experimentally determined masses matched (allowing one miscleavage and error within 50 ppm) to four Clp proteins, ClpS1, ClpP5, ClpP1, and ClpP6. The more intense matching peptides are indicated in the spectrum (*P1* for ClpP1, *S1* for ClpS1, etc). Trypsin autolysed peptides are indicated by *Tr*. Further details are described in Table I. The upper inset shows the spectrum at full scale, and the lower inset demonstrates the monoisotopic resolution of the spectrum. B, nano-electrospray MS spectrum in the m/z region from 400 to 1200 of the peptides generated by tryptic digestion of protein spot 2. A number of multiple charged peptide ions were selected for fragmentation (MS/MS), and eight of those ions matched to the 4 Clp proteins also observed in the MALDI-TOF MS spectrum. These precursor ions are indicated in the spectrum. The ESI-MS/MS spectrum for the double charged ion at m/z 848.4 is shown in C and the complete y series (y_1 – y_{15}) and interpreted amino acid sequence are indicated.

to ClpP1 (formerly pClpP) of *Arabidopsis*. This identification was further confirmed by another sequence tag from the triply charged peptide with m/z 1032.9 (see Table I).

Content of the Clp Complex and Discrepancies between Predicted and Expressed Proteins—The mass spectrometry analysis of the nine spots resulted in the positive identification of 10 different Clp proteins. In this section we will briefly describe

their identification. At this point it is important to comment on the nomenclature of the Clp proteins in higher plants. In agreement with a suggestion for nomenclature of chloroplast proteases (45), we have renamed a number of the Clp genes, based on the presence (or absence) of functional domains and conform the homology to *E. coli* and *Synechocystis PCC6803* Clp genes. This is summarized in Table I and in the section below. One

A ClpR3

MASCLQASMNLLPRSSSFSHPPLSSNSGRRNLKTFRYAFRAKASAKIMPPIPNDPFLSTLASIAANSPEKLLNRPVNADVPPYL
 DIFDPSQLMSSPAQVERSVAYNEHRPRTPPDLPMSLLDGRIVYIGMPT**DDIDFCFQ**LVPVAVTELVAELMYLQWLDPEKIYIYINST
 GTTR**DDGETVQMESEGFAYIDSLMQLKNEV**HTVCVGAAGIQAACLLLSAGTKGKRPFMMPHAK**GESEFDYFIFLIPVMQLOVAMIQQPRN**
FSSGLMPASDVLIRAKEVITNRDILVELLSKHTGNSVETVANVMRRPYMDAPKAKEFGVIDR**VSLGTVIFILWRGQEKIADVVPSE**
FDRNAGIKSVV

B ClpS1

MASYTVSFIPLTSLNPRIFVSRQNGSPSSSSRIPLTSSLLGKLLATQPSHRFCVPKLRCLTSASTVLNVPIAQENGSSDKIPKWSAR
 AIKSLMAGELBARKLYPSTGTEAILMGILVEGTSVAKFLRGNVTLFKVRDETLSSLGK**SDMYFFSPEHPPLTEPAQKAI**AJAWAIDEK
 NKSVDVDELTTAYLLLVWSQKDSAGRQILEKLGFNEDKAKEVEKS**MNEVDVLSFKKGG**

C ClpR4

MEVAATATSFTTLRARTSAIIPSSSTRNLRSKPRFSSSSSLRASLNGFLSPYTGGSISSDLCGAKLRAESLNPLNFSSSKPKRGVVTM
 VIPFSGKSAHEQPPDLASLYLAKNRIYVLGMSLVPSVTELLAEFLYLQYDEEKPIYLYINSTGTTKNGEKLYDTEAFAYIDVMGVV
 KPPIFTLCVGNAGEAALLLTAGAKGNRSALPSSSTIMIK**PIARFOGQATDVEIARKEIKHIKTEMVKLYSKHIGKSPQI**EADMRRPK
YFSPTAEVEYGIIDKVVYNERGSODRGVVSDDLKKAQLI

FIG. 3. Sequences of the ClpR3, ClpS1, and ClpR4. A, sequence of ClpR3 after correction of mis-assignment of predicted protein sequence from the genomic sequence (accession number 2342674). *Green* amino acid residues need to be inserted (from ESTs 8684305, 8685348, 8686535, 8680144, 8734821, and others), whereas the *red* areas need to be removed to arrive at the correct sequence. The matching MS/MS sequence tags are boxed. B, sequence of the ClpS1 protein after comparison of the predicted protein sequence from cDNAs (e.g. 8770652 and 8703777) and the predicted protein sequence from the genomic sequence (accession no. 7487579). *Light shaded* amino acid residues need to be added, to arrive at the correct sequence. The matching MS/MS sequence tag is boxed. C, corrected sequence of ClpR4 (accession number 7268455). *Light shaded* amino acid residues need to be added, to arrive at the correct sequence (from ESTs 8729730, 8691714, and others).

accession number per Clp protein is indicated in the main part of Table I; redundant entries are listed as footnotes, and discrepancies between the sequences are noted (Table I). The region between 20 and 30 kDa was divided completely in nine individual slices (spots). Since the spots are very close to each other and in some cases clearly overlapping, it is not surprising that indeed overlap in identities between neighboring spots was found (see below).

Spots 1 and 2 are the most intensely stained spots both with Silver and Coomassie staining (Fig. 1, B and C). Spot 1 contained the chloroplast-encoded subunit pClpP (here reassigned ClpP1), which was identified by six matching peptides MALDI-TOF MS with a total coverage of 33% and confirmed by one sequence tag. Spot 2 contained ClpP1 (formerly pClpP) as well as ClpP5 (formerly ClpP1) and a new Clp protein, here assigned ClpS1 (see “Discussion” and phylogenetic trees). These three proteins were identified by MALDI-TOF and confirmed by MS/MS (Fig. 2, A and B). In addition, we identified ClpP6 by one sequence tag, which was further strengthened by three matching peptide masses detected by MALDI-TOF MS.

Spots 3 and 4 contained ClpP5 with four matching peptides providing 26% of coverage and confirmed by five sequence tags, as well as ClpP2 (formerly annotated as Clp-like) identified by three sequence tags. ClpP2 was also identified in spots 5 by several sequence tags, as well as five matching peptides in MALDI TOF MS. Spot 6 contained ClpR2 (formerly annotated as nClpP2), as determined by eight matching peptides with 30% of coverage, and confirmed by three sequence tags. ClpR2 is interesting in that it has lost its catalytic site and was therefore renamed as ClpR (see below and “Discussion”).

Spots 7 and 8 have a partial overlap. Spot 7 contained mostly ClpP4 (formerly nClpP4), as was evident from eight matching peptides providing 56% of coverage and confirmed by four sequence tags. In addition, this protein spot also contains some ClpR3 (formerly nClpP8), as identified by MS/MS. Spot 8 contains ClpP4 with nine matching peptides with 32% of coverage, confirmed by one sequence tag, ClpR1 (formerly nClpP5) with nine matching peptides with 24% of coverage and ClpR3 (nClpP8) with nine matching peptides, 22% of coverage and confirmed by one sequence tag.

Finally, spot 9 had a partial overlap with spot 8 and contained ClpR3 (formerly nClpP8), ClpR1 (nClp5), and ClpP3. ClpP3 was identified by seven matching peptides in MALDI-

TOF MS, covering 34% of the predicted sequence, and was confirmed by a sequence tag in MS/MS.

During the identification process of the Clp genes, we encountered several discrepancies between the experimentally determined amino acid sequences and the predicted proteins from genomic sequence. Therefore, all predicted sequences were searched against the EST data bases to verify if these discrepancies result from incorrect intron/exon boundary prediction (since cDNAs have no introns) and also to identify possible discrepancies resulting from sequencing errors. Such mistakes in protein prediction were found in case of ClpR3, ClpS, as well as other Clp proteins not present in the complex (see “Discussion”), and after correction, the experimentally determined amino acid sequences matched to the predicted proteins (Fig. 3).

To arrive at the (most likely) correct protein sequence of ClpR3, three stretches of protein sequence were removed (resulting from incorrect intron boundary prediction and totaling 36 amino acid residues) and three stretches of sequence (totaling 75 amino acid residues) were inserted or added to the C terminus (Fig. 3A). Two of the experimentally determined sequences overlapped with the inserted amino acid sequence. After this correction, the predicted protein increased from 331 amino acid residues to 370 amino acid residues, thus adding ~4.3 kDa to the predicted molecular mass. In case of ClpS1, 14 amino acids were added to the C terminus of the predicted protein sequence (Fig. 3B).

Alignment and Functional Domains of the Clp Protein Identified in the 350-kDa Complex—To ensure and demonstrate that we were able to discriminate between the different ClpP and ClpR proteins, we aligned the ClpP1–6 and ClpR1–3 proteins and we have indicated the matching amino acid sequence tags determined by MS/MS (Fig. 4). Clearly, the MS/MS tags were unique for each protein, and, combined with the corresponding MALDI-TOF MS analysis, this shows that the proteins were positively identified with high certainty. The alignment also demonstrates the extent of conservation and the presence (ClpP1–6) or absence (ClpR1–3) of the catalytic triad (see “Discussion” and phylogenetic trees). ClpS1 (Fig. 3B) shows no strong homology and seems quite distant from the other identified Clp genes in the complex (see “Discussion” and phylogenetic trees) and was therefore not included in the alignment.

Localization Prediction, Transit Peptide, and Predicted Mo-

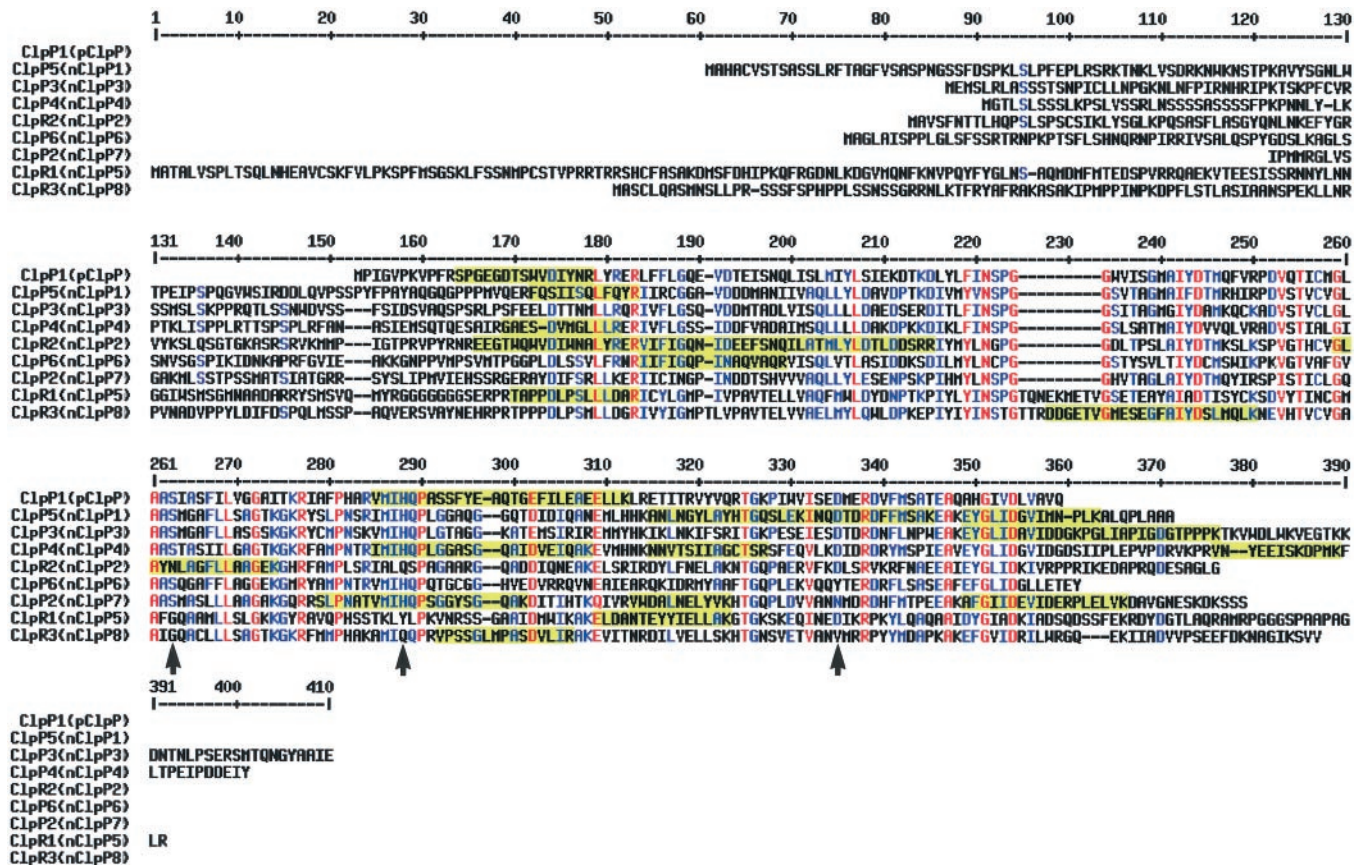


FIG. 4. Alignment of C_pP1–6 and Cl_pR1,2,3, identified in the 350-kDa complex, after correction for errors in gene prediction based on the genomic sequences. The completely conserved residues are colored in red, and the conserved residues that are similar are colored in blue. Arrows indicate the residues that have reported to form the active catalytic triad (but see “Discussion”). The matching MS/MS sequence tags are indicated in yellow. Multi-alignments were done using MultAlin.

lecular Mass—From the mass spectrometry analysis, we determined that the 350-kDa complex is a Clp protease complex with one chloroplast-encoded subunit (ClpP1) and nine nuclear-encoded subunits (five ClpP proteins, three ClpR proteins, and ClpS1). All the nuclear-encoded proteins are synthesized as precursor proteins with an N-terminal chloroplast signal peptide, which is cleaved off after import into the chloroplast. The chloroplast signal peptide has a number of distinct features but no conserved primary sequence, and can be used to predict chloroplast localization with moderate to good confidence (2, 40). To evaluate the chloroplast localization of the identified Clp proteins, the primary amino acid sequences were analyzed by the localization prediction programs TargetP and Predotar.

All identified Clp genes, except ClpP2, were predicted to be chloroplast localized by TargetP with high degree of confidence ($0.72 < p < 0.97$) (Table I). Since ClpP2 is clearly part of the same complex as the other subunits, it was surprising to find that this protein is predicted to be in the mitochondria. However, BLAST searching with this *Arabidopsis* protein identified several ESTs from rice, barley, and wheat; assembly of several overlapping ESTs for each of these plant species resulted in the identification of three Clp genes with high homology to *Arabidopsis* ClpP2 (Fig. 5). TargetP, predicted that all three proteins are localized in the chloroplast, whereas the *Arabidopsis* gene was predicted (but with ambiguity) to be in the mitochondria (Fig. 5). The alignment strongly suggests that the *Arabidopsis* protein sequence in the data base is truncated at the N terminus in the chloroplast transit peptide (~34 amino acids are missing) (Fig. 5). We could not (yet) identify an EST covering the N terminus and the genome sequence allows for several possible initiat-

ing methionines. With the search program Predotar, ClpP2 of wheat and barley were also predicted to be localized in the chloroplast, whereas the rice protein was predicted to be neither in the chloroplast nor in the mitochondrion.

After the corrections of the predicted protein sequence from genomic nucleotide sequence data and removal of the predicted transit peptide, most of the predicted molecular masses fitted with the position on the SDS-PAGE gel. This further strengthened all identifications as well as the chloroplast localization. However, the predicted size of ClpR1, which was identified by a strong set of MS data, seems to be several kDa larger than expected from the SDS-PAGE gel and warrants further analysis. Additionally, the sequence alignment in Fig. 4 indicates that ClpR1 has an extended N-terminal region, as compared with the other ClpP and ClpR proteins, but no cDNA sequence data are available to verify the N terminus of ClpR1.

DISCUSSION

Isolation and Identification of a Chloroplast Clp Complex—In this study we have identified a chloroplast Clp complex of *Arabidopsis thaliana*, using Blue-Native gel electrophoresis and mass spectrometry. The high resolution in the first dimension of gel electrophoresis allowed us to separate the Clp complex, with estimated molecular mass of 330–370 kDa, from the 400-kDa CF1 complex. The Clp complex was coisolated with thylakoid membranes, and subsequently released by Yeda press treatment. When the same procedure was followed in absence of MgCl₂, the Clp complex could not be detected (data not shown), indicating that Mg²⁺ was either required to stabilize the Clp complex or determines the affinity to the membrane. An earlier study showed that ClpP1 and ClpC are

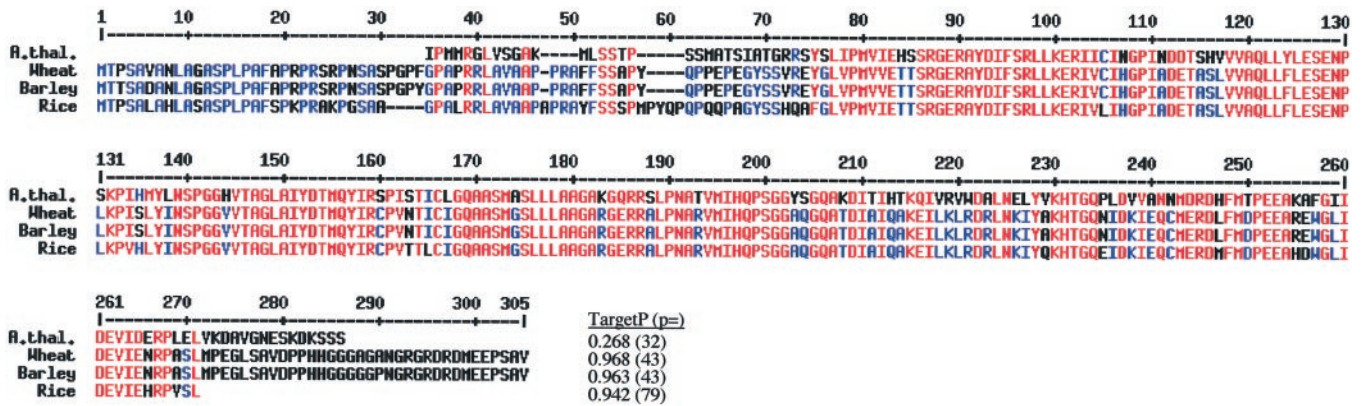


FIG. 5. Alignment of the predicted *Arabidopsis* protein sequence of Clp2 with assembled translated EST sequences from wheat (gi 9360943, 8860392, 9423733), barley (gi 8901631, 9465888, 9823207), and rice (gi 8956172, 2798641). The prediction by TargetP for chloroplast localization, as well as the most likely primary cleavage site of the chloroplast transit peptide, is indicated. Multi-alignments were done using MultAlin. Color coding is as explained in the legend to Fig. 4.

localized in the chloroplast stroma (16), and it is therefore likely that only a subfraction of these proteins bind (as a complex) to the thylakoid membranes. There is very little information on the precise localization and substrates of the active Clp complex. Interestingly, proteolytic disposal of fully or partially assembled cytochrome *b₆* of the thylakoid membrane of green algae *Chlamydomonas reinhardtii*, is controlled by the Clp protease (25). This possibly indicates that the ClpP complex can operate in close vicinity of the thylakoid surface.

SDS-PAGE of the complex, followed by Silver staining showed that the complex falls into at least seven different protein spots in the range of 20–30 kDa. No spots outside this region were detected after high sensitivity silver staining. We analyzed the complete region from 20 to 30 kDa by mass spectrometry and we identified 10 different Clp proteins (Table I), indicating that the chloroplast ClpP core complex is surprisingly complicated, as compared with related complexes in photosynthetic bacteria and other prokaryotes, such as *E. coli*. Expression and accumulation of several of the identified *Arabidopsis* Clp proteins have not been shown before. It is rather unlikely that the complex contained significant levels of any additional Clp related proteins, since we carried out such detailed mass spectrometry analysis and since the *Arabidopsis* genome is now completely sequenced.

Corrections of Gene Assignment—The genomic sequences of all nuclear-encoded Clp proteins in the isolated Clp complex are present in the data base. ClpP2, ClpP5, ClpP6, ClpR1, and ClpR3 are located on chromosome I, ClpS1 is located on chromosome IV, and ClpP2 and ClpP4 are located on chromosome V. In addition to the earlier discussed redundancy in data base entries (see Table I), we noticed several significant discrepancies between the protein sequence that we experimentally determined by mass spectrometry and the predicted protein sequence (from genomic sequence). This triggered us to compare the predicted proteins from genomic data with predicted proteins from EST sequences. For ClpR3 we could thus determine that in total 36 amino acid residues needed to be removed and 75 amino acid residues needed to be inserted or added, whereas, in the case of ClpS1, 14 amino acids needed to be added at the C terminus (Fig. 3, A and B). After these corrections, the experimentally determined sequence tags matched with the predicted sequence. Alignment of the predicted sequence of *Arabidopsis* ClpP2 with predicted sequences of wheat, barley, and rice (assembled from cDNAs) showed that the N terminus of the predicted *Arabidopsis* gene was probably missing the first 34 amino acids (Fig. 5).

Phylogenetic Trees, Sequence Analysis, and Localization—As

an attempt to better understand the complexity of the identified Clp complex, we investigated the relationship between the identified (and corrected) ClpP, ClpR, and ClpS proteins and the members of the Clp/HSP100 family identified in *Arabidopsis* (Fig. 6A). A number of additional genes were also identified (here assigned ClpR4, CpS2, ClpF, ClpZ, and Clp-like 1 and 2), that were related to ClpP/R/S and/or to ClpB, -C, -D, and -X. We have carefully verified that all protein sequences used for this analysis were truly derived from different genes, and, in case there were two or more entries for the same protein, we have selected the one that looked most complete. The ClpR4 sequence was corrected using overlapping EST sequences (Fig. 3C). The annotated C terminus of ClpS2 needed to be corrected by removing the annotated 40 C-terminal amino acids by the following sequence: ELESFASESGLFDE.

The relationships between the different Clp proteins was studied by phylogenetic tree analysis. Because some of the sequences were extremely divergent, establishing a reliable phylogenetic relationship was not straightforward (see “Experimental Procedures” for details). A *black box* is placed at the center of the tree where homology cannot be established based upon sequence alone. The relationships between the ClpB, C and D proteins are overall quite ambiguous, and part of this clade is therefore depicted as a star cluster. ClpF has been named here and shows strong homology to the N-terminal half of ClpC, including the N-terminal “signature” (see Ref. 10). ClpZ has a strong homology with the C terminus of ClpB1–4, but otherwise no signature can be recognized. ClpS1,S2, as well as ClpX1,X2 form two separate clades, whereas the two Clp-like proteins are not directly related. These Clp-like 1 and 2 proteins are neither clearly chaperones-like Clp proteins nor real ClpP/R proteins, and we have just named them Clp-like1 and -2. We would like to note that, although ClpC in *Arabidopsis* and other higher plants belongs to class I of Clp/HSP100 family, it has just one NBD (10). The ClpP/R clade forms a distinct group.

The 10 different proteins that we identified in the 350-kDa complex (boxed in red in Fig. 6A) are all clustered in the CpP/R clade, with the exception of ClpS1. Interestingly, ClpS1,2 seem to be plant-specific. ClpS1,2 have no NBDs but share homology with the N terminus of ClpC. ClpS1,2 also share weak similarity with the ClpP proteins, but the catalytic triad (see below) is not conserved. Neither ClpS2 nor ClpR4 was identified in the complex, although both are predicted to be localized in the chloroplast.

All Clp protein sequences were evaluated for possible chloroplast or mitochondrial prediction. Sixteen of the Clp proteins

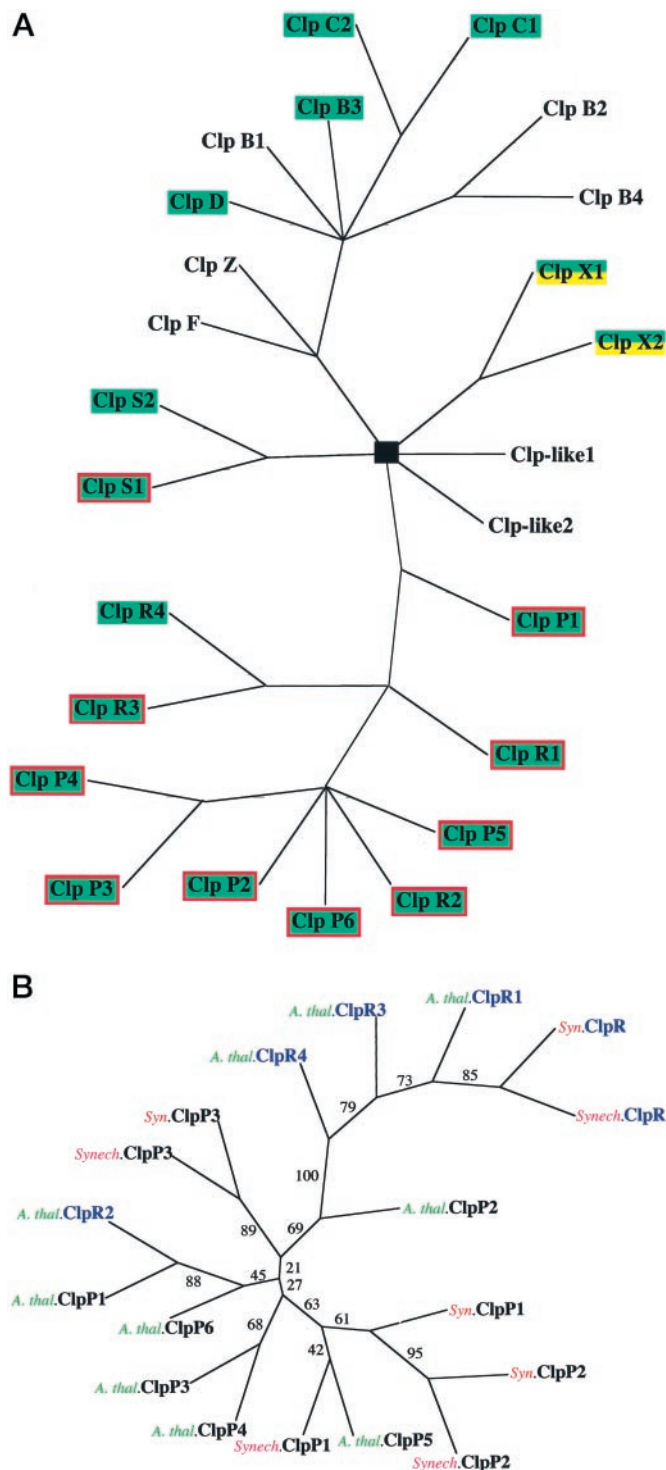


FIG. 6. Phylogenetic tree analysis of the Clp proteins of *Arabidopsis thaliana*, *Synechocystis*, and *Synechococcus*. ClpR proteins 1–3, ClpS1,2 and the chaperone-like ClpB–D family, as well as ClpX1,2. Four additional Clp genes are also included. Accession numbers for the ClpP1–6, ClpR1–4 and ClpS1 are listed in Table I and Fig. 3. The proteins that are predicted to be localized in the chloroplast are in *green*. The prediction for the location of the ClpX proteins is ambiguous and could be the chloroplasts or mitochondria or both; they are colored in *yellow/green*. Proteins that have been identified in the 350-kDa Clp complex are boxed in *red*. *A*, phylogenetic tree analysis of the ClpP proteins. A box is placed at the center of the tree, where homology cannot be established based upon sequence alone. Accession numbers for the other proteins are listed, and the chromosome location is indicated in *roman numbers*: 537446 (*I*; ClpB1), 6623879 (*II*; ClpB2), 9755800 (*V*; ClpB3), 7435720 (*IV*; ClpB4), 2921158 (*V*; ClpC1), 5360574 (*III*; ClpC2), 1169544 (*V*; ClpD), 6996248 (*III*; ClpF), 7267907 (*IV*; ClpS2), 9759182 (*V*; ClpX1), 8978260 (*V*; ClpX2), 8777573 (*III*; ClpZ),

are predicted to be localized in the chloroplast and they are indicated in *green* (Fig. 6A). The prediction for ClpX1,2 is ambiguous; TargetP predicts that ClpX1 and ClpX2 are respectively localized in chloroplasts and in the mitochondria (in confidence class rc2 and rc1, respectively; see Ref. 2), whereas Predotar makes the opposite prediction. It has been mentioned that at least one of the ClpX proteins is located in plant mitochondria (14), but experimental evidence has not been published. ClpX homologues are present in mitochondria in mammalian cells (41) and yeast (assigned Mcx1p) (42).

Amino acid residues Ser-97, His-122, and Asp-171 have been shown to be essential for the catalytic activity of *E. coli* ClpP. In *Arabidopsis* ClpP1, -3, -4, and -5 possess this triad, whereas Asp-171 is replaced by an Asn in ClpP2 and replaced by a Tyr Y in ClpP6. The ClpP2 homologues in wheat, barley, and rice have a Cys at this position (see Fig. 5). It has been reported that, in some cases, Asp-171 can be replaced by a different amino acid, with the protein still being functional (43).² Careful analysis of the multi-aligned *Arabidopsis* Clp proteins revealed that the only amino acid strictly correlated with the presence of Ser-97 and His-122 in all the ClpP genes is Asp-174. It is therefore tempting to suggest that Asp-174 is the third component of the catalytic triad in plants, rather than Asp-171.

Chloroplasts are thought to originate from cyanobacteria. The genome of the cyanobacterium *Synechocystis PCC6803* has been completely sequenced and contains three ClpP genes and one ClpR gene (previously named ClpP4). In addition, three ClpP genes and one ClpR gene were also discovered in the cyanobacterium *Synechococcus* (44). In an attempt to understand if the ClpR genes in chloroplasts derived directly from the invading prokaryotic ancestor or from the ClpP genes after the endosymbiotic event, a second tree was built using the ClpP/R protein sequences from *Arabidopsis*, *Synechocystis*, and *Synechococcus* (Fig. 6B), with the same methodology as for Fig. 6A. The tree shows that *Arabidopsis* ClpR1, -3, and -4 proteins form a clade together with the two bacterial ClpR proteins, suggesting that ClpR was probably already present prior to endosymbiosis. Interestingly, *Arabidopsis* ClpR2 is more closely related to several of the *Arabidopsis* ClpP proteins, suggesting that ClpR2 may be derived from a ClpP gene within the evolving plant, either in the nuclear or the chloroplast genome.

Finally, it is intriguing that ClpP1 is retained in the plastid genome in all sequenced higher plant species and green algae, even in the nonphotosynthetic plant *Epifagus* in which only 42 genes are left in the plastid genome, whereas the other (24) identified Clp genes are located on the nuclear chromosomes. Deletion of the chloroplast-encoded ClpP gene is most likely to be lethal in *C. reinhardtii* (26). One explanation could be that the ClpP transcript plays a role in assembly of the Clp core complex.

Stoichiometry of the Clp Proteins in the Isolated Complex and Their Physiological Role—Given the strong conservation of Clp proteins between bacteria, algae, plants, and mammals, and reported structures of *E. coli* and human Clp complexes, it is highly likely that the isolated chloroplast 350-kDa ClpP/R/S complex consists of 2 heptameric rings. This would lead to an average molecular mass of 25 kDa for the 14 subunits. Since all identified Clp proteins

² A. Wlodower, personal communication.

6728865 (*I*; Clp-like1), 6630747 (*III*; Clp-like2). *B*, neighbor-joining tree built in PHYLIP, with indicated bootstrap values for the ClpP/R proteins in *Arabidopsis* and the cyanobacteria *Synechococcus* (*Synech.*) and *Synechocystis* (*Syn.*). Accession numbers for the *Arabidopsis* ClpP/R proteins are listed in Table I and Fig. 3. Accession numbers for the *Synechococcus* proteins are 755165 (ClpP1), 2351823 (ClpP2), (7106112) ClpP3, and 7264063 (ClpR), and for *Synechocystis PCC6803* are 1705930 (Clp1), 2493737 (Clp2), 3023511 (Clp3), and 1653655 (Clp4).

in the complex fall in the range of 22–29 kDa, this seems quite reasonable. We believe that the 350-kDa complex does not contain significant levels of additional Clp related proteins.

Although the mass resolution of the Blue-Native PAGE was good (~25 kDa), we cannot exclude a superposition of several of Clp complexes of different composition/stoichiometry with approximately the same molecular mass. Since the molecular masses of the different ClpP/R/S proteins are within a narrow range, it is possible to form a tetradecamer with different combinations of isoforms.

The ClpR proteins are homologous to the ClpP proteins, and it is likely that a mixture of ClpR and ClpP proteins are able to form a heptameric ring. The ClpR proteins have probably no proteolytic activity, and such a heptameric ring with mixed isomers might thus have a lower overall proteolytic activity compared with heptamers made up of active ClpP proteins. The presence of ClpS1 in the complex is intriguing since ClpS1 is quite distant from ClpP/R and since ClpS1 has also features of the chaperone family. We speculate that the ClpS1 protein is positioned at the axial opening of the complex, rather than being part of the hexameric rings, possibly fulfilling a role of assisting proteins in and out of the Clp complex. The mass spectrometry analysis suggests that ClpP5 is more abundant than ClpS1, which would be consistent with the idea that ClpS is a regulatory subunit and is no part of the heptameric rings.

We did not identify ClpS2 and ClpR4 in the isolated complex, although transit peptide analysis indicated a chloroplast localization. Since EST sequencing demonstrated that these genes are transcribed, it is possibly that they are expressed only at very low (undetectable) levels or are expressed under particular stress conditions or during specific developmental stages of the chloroplast. We used leaves of unstressed fully developed *Arabidopsis* plants in the vegetative stage, and it is likely that the content of the Clp complex isolated in this study is a reflection of this leaf stage and growth conditions.

Conclusions and Future Studies—The combination of high resolution nondenaturing electrophoresis and modern mass spectrometry, together with the nearly completely sequenced *Arabidopsis* genome, has allowed rapid identification of chloroplast localized multi-subunit ClpP/R/S complex. Identification of a larger Clp complex that includes members of the Clp/HSP100 family is a next step that should be within reach. Study of the protein composition and structure of the Clp core complex during different stresses and development should provide insight to understand why and how the Clp complex has evolved toward such a tremendous complexity in the chloroplasts of photosynthetic organisms. The combination of Blue-Native PAGE and mass spectrometry is a very powerful tool for study of protein-protein interactions, as demonstrated in this study, and can be applied to any well sequenced organism. Finally, experimental verification of protein predictions based of the sequenced genome of *Arabidopsis thaliana* is needed and can be effectively carried out using proteomics tools.

Acknowledgments—We thank Soren Andersen, Kate Rafn, Ole N. Jensen, and Dario Kalume in Odense for their advice and assistance with mass spectrometry and members of the laboratory of B. Andersson

for providing the Yeda press.

REFERENCES

1. Abdallah, F., Salamini, F., and Leister, D. (2000) *Trends Plant Sci.* **5**, 141–142
2. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) *J. Mol. Biol.* **300**, 1005–1016
3. van Wijk, K. J. (2001) in *Regulatory Aspects of Photosynthesis: Advances in Photosynthesis*. (Andersson, B., and Aro, E.-M., eds) Vol. 8, Kluwer Academic Press, Dordrecht, The Netherlands, in press
4. Schagger, H., Cramer, W. A., and von Jagow, G. (1994) *Anal. Biochem.* **217**, 220–230
5. Jensen, O. N., Larsen, M. R., and Roepstorff, P. (1998) *Proteins*, Suppl. 2, 74–89
6. Pandey, A., and Mann, M. (2000) *Nature* **405**, 837–846
7. Katayama, Y., Gottesman, S., Pumphrey, J., Rudikoff, S., Clark, W. P., and Maurizi, M. R. (1988) *J. Biol. Chem.* **263**, 15226–15236
8. Katayama-Fujimura, Y., Gottesman, S., and Maurizi, M. R. (1987) *J. Biol. Chem.* **262**, 4477–4485
9. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C., and Maurizi, M. R. (1998) *J. Biol. Chem.* **273**, 12476–12481
10. Schirmer, D. C., Glover, J. R., Singer, M. A., and Lindquist, S. (1996) *Trends Biochem. Sci.* **21**, 289–296
11. Wang, J., Hartling, J. A., and Flanagan, J. M. (1997) *Cell* **91**, 447–456
12. Maurizi, M. R., Thompson, M. W., Singh, S. K., and Kim, S. H. (1994) *Methods Enzymol.* **244**, 314–331
13. Missiakos, D., Schwager, F., Betton, J. M., Georgopoulos, C., and Raina, S. (1996) *EMBO J.* **15**, 6899–6909
14. Adam, Z. (2000) *Biochimie* **82**, 647–654
15. Clarke, A. K. (1999) *Ann. Botany* **83**, 593–599
16. Shanklin, J., DeWitt, N. D., and Flanagan, J. M. (1995) *Plant Cell* **7**, 1713–1722
17. Nakabayashi, K., Ito, M., Kiyosue, T., Shinozaki, K., and Watanabe, A. (1999) *Plant Cell Physiol.* **40**, 504–514
18. Weaver, L. M., Froehlich, J. E., and Amasino, R. M. (1999) *Plant Physiol.* **119**, 1209–1216
19. Moore, T., and Keegstra, K. (1993) *Plant Mol. Biol.* **21**, 525–537
20. Miller, J. D., Artega, R. N., and Pell, E. J. (1999) *Plant Physiol.* **120**, 1015–1024
21. Desimone, M., Weiss-Wichert, C., Wagner, E., Altenfeld, U., and Johanninger, U. (1997) *Bot. Acta* **110**, 234–239
22. Sokolenko, A., Lerbs-Mache, S., Altschmied, L., and Herrmann, R. G. (1998) *Planta* **207**, 286–295
23. Liu, X.-Q., and Jagendorf, A. T. (1984) *FEBS Lett.* **166**, 248–252
24. Halperin, T., and Adam, Z. (1996) *Plant Mol. Biol.* **30**, 925–933
25. Majeran, W., Wollman, F. A., and Vallon, O. (2000) *Plant Cell* **12**, 137–150
26. Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M., and Liu, X.-Q. (1994) *Mol. Gen. Genet.* **244**, 151–159
27. Damerau, K., and St. John, A. C. (1993) *J. Bacteriol.* **175**, 53–63
28. Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., and Gottesman, S. (1990) *J. Biol. Chem.* **265**, 12536–12545
29. Schagger, H., and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
30. Rabilloud, T., Vuillard, L., Gilly, C., and Lawrence, J. J. (1994) *Cell Mol. Biol.* **40**, 57–75
31. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal. Chem.* **68**, 850–858
32. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., and Roepstorff, P. (1999) *J. Mass Spectrom.* **34**, 105–116
33. Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K. J. (2000) *Plant Cell* **12**, 319–342
34. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) *Nature* **379**, 466–469
35. Benner, S. A., Chamberlin, S. G., Liberles, D. A., Govindarajan, S., and Knecht, L. (2000) *Res. Microbiol.* **151**, 97–106
36. Yona, G., Linial, N., and Linial, M. (1999) *Proteins* **37**, 360–378
37. Felsenstein, J. (1989) *Cladistics* **5**, 164–166
38. Keegstra, K., and Cline, K. (1999) *Plant Cell* **11**, 557–570
39. Burlingame, A. L., Boyd, R. K., and Gaskell, S. J. (1998) *Anal. Chem.* **70**, 647R–716R
40. van Wijk, K. J. (2000) *Trends Plant Sci.* **5**, 420–425
41. Santagata, S., Bhattacharyya, D., Wang, F. H., Singha, N., Hodtsev, A., and Spanopoulou, E. (1999) *J. Biol. Chem.* **274**, 16311–16319
42. van Dyck, L., Dembowski, M., Neupert, W., and Langer, T. (1998) *FEBS Lett.* **438**, 250–254
43. Dodson, G., and Wlodower, A. (1998) *Trends Biochem. Sci.* **23**, 347–352
44. Clarke, A. K., Schelin, J., and Porankiewicz, J. (1998) *Plant Mol. Biol.* **37**, 791–801
45. Adam, Z., Adamska, I., Nakabayashi, K., Rodermer, S. R., Shinozaki, K., and Clarke, A. K. (2001) *Plant Physiol.*, **125**, 1912–1918