Tetradecameric Clp protease core complexes in non-photosynthetic plastids of roots, flower petals, and in chloroplasts of leaves of Arabidopsis thaliana were purified based on native mass and isoelectric point and identified by mass spectrometry. The stoichiometry between the subunits was determined. The protease complex consisted of one to three copies of five different serine-type protease Clp proteins (ClpP1,3–6) and four non-proteolytic ClpR proteins (ClpR1–4). Three-dimensional homology modeling showed that the ClpP/R proteins fit well together in a tetradecameric complex and also indicated unique contributions for each protein. Lateral exit gates for proteolysis products are proposed. In addition, ClpS1,2, unique to land plants, tightly interacted with this core complex, with one copy of each per complex. The three-dimensional modeling show that they do fit well on the axial sites of the ClpPR cores. In contrast to plastids, plant mitochondria contained a single ~320-kDa homo-tetradecameric ClpP2 complex, without association of ClpR or ClpS proteins. It is surprising that the Clp core composition appears identical in all three plastid types, despite the remarkable differences in plastid proteome composition. This suggests that regulation of plastid proteolysis by the Clp machinery is not through differential regulation of ClpP/R/S gene expression, but rather through substrate recognition mechanisms and regulated interaction of chaperone-like molecules (ClpS1,2 and others) to the ClpP/R core.

Plastids are essential organelles of prokaryotic origin that are present in every plant cell and differentiate from proplastids into non-photosynthetic plastids in roots and flowers and photosynthetic plastids in leaves and stems. Plastids are responsible for synthesis of key molecules required for the architecture and functions of plant cells.

To maintain a correct stoichiometry between different proteins and pathways, to remove and recycle damaged or misfolded proteins, and to control gene expression by proteolysis of transcription or translation factors, different proteolytic systems are present in the plastid. Members of at least five protease families are present in plastids, but their structures, functions, substrates, and biological importance are poorly understood.

A very prominent group of proteases in plants is the Clp protease family. Our latest analysis of the Arabidopsis thaliana nuclear genome indicates the presence of at least 26 Clp-related genes, with 15 genes encoding for plastid-localized proteins (3) (Fig. 1). An additional gene, ClpP1, is located on the plastid genome and is essential in tobacco (4, 5) and Chlamydomonas reinhardtii (6). Reduced expression of the ClpP1 gene in C. reinhardtii suggested that ClpP1 is involved in degradation of photosynthetic complexes (7). Transcript analysis suggested that in plants the Clp genes are constitutively expressed, with only minor altered gene expression under specific stress conditions or during senescence (8–10). In the photosynthetic cyanobacterium Synechococcus, the ClpP1–3, ClpR, ClpX, and ClpC genes are essential for autotrophic growth (11, 12).

Clp proteins (for caseinolytic protease) were first identified in Escherichia coli (13). In E. coli, ClpP complexes were shown to be ATP-dependent serine-type proteases, consisting of a catalytic core, flanked at one or both ends by a hexamer of regulatory ClpA/X chaperone subunits, and belonging to the family of ATP-dependent HSP100 proteins (14). This catalytic core is composed of two homo-heptameric rings of ClpP subunits, all encoded by a single gene. The diameter of the cavity is 50 Å, and the diameter of entrances at either end of the complex was reported as 10 Å (15). This suggested that only peptides and proteins in an unfolded state can enter the cavity, which is in agreement with biochemical data showing that the ATPase and ATP hydrolysis are required for the degradation of folded proteins (16). However, E. coli ClpXP can accommodate 2 or 3 polypeptide chains at the same time, leading to an estimation of the opening of 20–25 Å wide (17). Clp complexes in bacteria can also be formed by ClpQ,Y (HslV and HsIU) proteases, which are unrelated to ClpP (18). No ClpQ,Y homologues have been found in any plant species.

The functional role of Clp complexes in E. coli and other bacteria, such as Bacillus subtilis, involves (i) degradation of mis-folded proteins, (ii) stress responses, and (iii) gene regula-
tion by proteolysis of transcription factors. Recognition of proteins by the Clp machinery in bacteria occurs via N- or C-terminal tags or via internal domains (see Refs. 19–21 and for discussion see Refs. 22 and 23).

In this study, we have investigated the Clp machinery in non-photosynthetic plastids of roots and petals, in chloroplasts of leaves, as well as in mitochondria. All three plastid types were found to contain a single tetradecameric Clp protease core complex (~325 kDa and pI ~5.0) consisting of one to three copies of five different serine-type protease Clp proteins (ClpP1,3–6) and four non-proteolytic ClpR proteins (ClpR1–4). In addition, two chaperone-like proteins, ClpS1,2, unique to land plants, were found to interact tightly with this core complex. In contrast, mitochondria contained a single homo-tetradecameric ClpP2 complex, without association of ClpR or ClpS proteins.

Large scale crystallization projects and three-dimensional homology modeling studies have shown that the structure of a protein can be predicted accurately from the x-ray crystal structure of a homologous protein. This removes the need to crystallize each protein or protein complex to know their structure (24, 25). Given the high level of sequence identity between ClpP sequences, we decided to build ClpP,R models using homology modeling studies have shown that the structure of a homologous protein. This removes the need to crystallize each protein or protein complex to know their structure. In our ClpP,R models are expected from the Protein Data Bank (www.rcsb.org/pdb/), as well as via cbus.tc.cornell.edu/vanwijk/.

**EXPERIMENTAL PROCEDURES**

**Plant Material and Protein Preparations—** *A. thaliana* (Col 0) was grown under 10-h light/14-h dark cycles at 25/17 °C, and leaves were collected from mature rosettes. Flowers were collected from *Brassica oleracea* grown with supplemental lightning. *Brassica rapa* and potato tubers were purchased at local stores. Thylakoids and peripheral complexes from *A. thaliana* were purified according to Ref. 26, with minor modifications to improve thylakoid purity. Plant material containing modifications to improve thylakoid purity. Plant material was then loaded at 100 × g and further purified on 40–85% Percoll cushions (Percoll in 0.8% Ficoll, 1.8% polyethylene glycol, and 0.6% bovine serum albumin) by a 10-min spin at 3750 × g and one additional wash in the grinding medium. Plastids were subsequently lysed in 10 mM Hepes, pH 8.0, 330 mM sorbitol and filtered through nylon mesh. The crude plastids were then collected by a 2-min spin at 1100 × g and further purified on 40–85% Percoll cushions (Percoll in 0.8% Ficoll, 1.8% polyethylene glycol, and 0.6% bovine serum albumin) by a 10-min spin at 3750 × g and one additional wash in the grinding medium. Plastids were subsequently lysed in 10 mM Hepes, pH 8.0, 5 mM MgCl2 with a mixture of protease inhibitors under mild mechanical disruption. The stromal content was collected, concentrated, and immediately used for two-dimensional analysis. Intact potato mitochondria were purified according to Ref. 27 with minor modifications.

**Protein Separation by Colorless Native Gels (CN-PAGE) and Native IEF**—Concentrated protein samples were supplemented with aminocaproic acid, BisTris/HCl, pH 7.0, and glycerol to a final concentration of 5 and 50 mM and 20%, respectively. Samples were immediately loaded on native gradient gels (6–15% acrylamide), according to Ref. 28. After separation, gels were cut, and the proteins were destained, reduced, alkylated, and run on Tricine-SDS-PAGE acrylamide gradient gels, according to Refs. 3 and 29. For native IEF, chloroplast stroma (10–15 mg) was loaded on 2.2-ml (0.1–1.1 ml) sucrose gradients, spun for 2.5 h at 259,000 × g, and collected in 0.2-ml fractions. The fractions were then loaded on IEF tube gels with pI range of 3 to 10 and 4 to 5, respectively as described previously (30). Proteins were focused for 20 kV-h, subsequently solubilized, reduced, and alkylated as described earlier (3). The gels were then loaded on Tricine-SDS-PAGE gels, as described for CN-PAGE above, and proteins were separated.

**Protein Identification by MS and Bioinformatics**—Stained protein spots were excised, washed, digested with modified trypsin, and peptides extracted automatically (Promega and Proseq, Genomic Solutions, Ann Arbor, MI), and peptides were applied to the MALDI-TOF target plates as described previously (1, 26). Automated MS or MS/MS analyses were carried out as follows. A Voyager-DE-STR (Perseptive Biosystems) generated the MALDI-TOF-MS spectra in reflectron mode, which were then annotated and internally calibrated. For nano-LC-ESI-MS/MS, a quadrupole-time of flight (micromass) was used. Separation of peptide mixtures was achieved by using a μ-Guard Pre-Column (LC Packings; inner diameter 300 μm, 1 mm long) followed by a PepMap C18 Reverse Phase Nano Column (LC Packings; inner diameter 180 μm, 15 cm long). Peptides were eluted by running a gradient

| Name       | AGI locus | Predicted location | Domains          | Comment |
|------------|-----------|--------------------|------------------|---------|
| ClpP1⁴     | Atcg00460 | P                  | Catalytic triad  | Plastid core |
| ClpP2      | Atcg23140 | P/M                | Catalytic triad  | Mitochondrial core |
| ClpP3      | Atkg8670  | P                  | Catalytic triad  | Plastid core |
| ClpP4      | Atcg45590 | P                  | Catalytic triad  | Plastid core |
| ClpP5      | Atkg02560 | P                  | Catalytic triad  | Plastid core |
| ClpP6      | Atkg11750 | P                  | Catalytic triad  | Plastid core |
| ClpR1      | Atkg19970 | P                  | No catalytic triad | Plastid core |
| ClpR2      | Atkg12410 | P                  | No catalytic triad | Plastid core |
| ClpR3      | Atkg09130 | P                  | No catalytic triad | Plastid core |
| ClpR4      | Atkg17040 | P                  | No catalytic triad | Plastid core |
| ClpS1      | Atkg23370 | P                  | No NBD           | Plastid core, I/LGF motif |
| ClpS2      | Atkg12060 | P                  | No NBD           | Plastid core, I/LGF motif |
| ClpX1      | Atkg53350 | P/M                | 1 NBD            | Chaperone, I/LGF motif |
| ClpX2      | Atkg49840 | P/M                | 1 NBD            | Chaperone, I/LGF motif |
| ClpX3      | Atkg33360 | M                  | 1 NBD            | Chaperone, I/LGF motif |
| ClpD       | Atkg1070  | P                  | 2 NBDs           | Chaperone, I/LGF motif |
| ClpC1      | Atkg0920  | P                  | 2 NBD            | Chaperone, I/LGF motif |
| ClpC2      | Atkg48870 | P                  | 2 NBD            | Chaperone, I/LGF motif |
| ClpB3      | Atkg15450 | P                  | 2 NBDs           | Chaperone, no I/LGF motif |
| ClpT       | Atkg8860  | P                  | Homologue of E. coli modulator |

⁴ Prediction by TargetP.
⁵ Plastid gene.
⁶ One of the nucleotide binding domain (NBD) is somewhat unusual (see Ref. 3).
Fig. 1. Summarizing overview of the Clp complexes in A. thaliana. A. Clp gene locations on the five A. thaliana chromosomes. ClpP1 is located on the plastid genome. Proteins predicted to the chloroplast are in green and those in mitochondrial and plastid in blue. The localization predictions of ClpP2, ClpX1, and -X3 are ambiguous. B, sequence alignment around the (I/L)GF motif for each of the groups of plastid and mitochondrial chaperones. The motif is implied in interaction between chaperones and core in case of ClpX and ClpA homologues.

Criteria for positive identification of proteins by peptide mass fingerprinting were at least 5 matching unique peptides (with mass accuracy <50 ppm, but typically <10 ppm), allowing no missed cleavage or modifications, and requiring at least 15% coverage of the predicted protein sequence. Criteria for nano-LC-ESI-MS/MS were at least two matching sequence tags of high quality (clear Y-ion series and several complementary B ions; manually verified). In case of identification of closely related paralogues or orthologues, it was verified that these sequences were unique. The proteome data will become available via cbsu.tc.cornell.edu/and as Tables 1–4 in the Supplemental Material.

ClpP2::GFP Fusion and Transient Expression Assay—The ClpP2 cDNA was amplified from an A. thaliana cDNA library (CD4–7; Arabidopsis Biological Resource Center). The cDNA was then cloned as an Ncol/SpeI fragment into p1302, a pCAMBIA vector (www.cambia.org), containing a mGFP5 expression cassette driven by a CaMV35S promoter. The resulting plasmid was sequenced and used for particle bombardment of Nicotiana tabacum cv. petit havana leaves on agar plates. Leaves were bombarded twice with plasmid-coated 1.1-μm tungsten particles using a model PDS 1000/He Biolistic Particle Delivery System and were kept in the dark for 3 days before imaging. Confocal laser scanning microscopy was performed on an MRC600 attached to a Zeiss Axiovert10 and COMOS program. Images were colored and projected using the Confocal Assistant 4.02 program. Equipment and software were from Bio-Rad.

Homology Modeling and Threading—Reliable pairwise alignments were obtained by BLAST. Three-dimensional models were generated by comparative protein modeling through satisfaction of spatial restraints with the program MODELLER (31), minimizing the violations of distance and dihedral angle restraints from the template structure (i.e., the x-ray crystal structure of eClpP, 1TYF). The main chain atoms of the models are expected to be within 0.4 Å of the equivalent atoms in the template structure in the conserved regions. Excluding the C-terminal regions, the Cα r.m.s.d. in the ClpP,R models are expected to be on average around 1.5 Å, with 70% of the side chain positions modeled correctly (32). The models of residues in the C-terminal part of ClpP,R proteins as well as those in the insertion loop (L1) of ClpR1,3,4 are built based on the x-ray crystal structures of structural neighbors (Protein Data Bank codes 1NZY and 1HNO) of the structure of eClpP, 1TYF. These parts of the models are expected to be within a 6-Å r.m.s.d. of the real structure.

By using the threading program 3DPSSM (33) and MODELLER, we built the models for ClpS1,2, respectively, based on the N-terminal domain. All models were validated using the WHAT_CHECK verification routines (34) from the program WHAT_CHECK (35). The coordinates of our Clp models can be downloaded from the Protein Data Bank (www.rcsb.org/pdb/), as well as via cbsu.tc.cornell.edu/and as Tables 1–4 in the Supplemental Material.

RESULTS

The Complexity of the Clp Gene Family in A. thaliana—Our latest analysis of the sequenced A. thaliana nuclear genome indicates the presence of 26 Clp-related genes, with 19 genes predicted to encode for plastid- or mitochondrial proteins (Table I and Fig. 1A). An additional ClpP gene (P1) is located on the plastid genome. The subcellular localizations of 16 of these organellar proteins have been experimentally determined in this study and/or in earlier studies (2, 3, 10). Recently, a new protein, ClpS, was identified in E. coli. This bacterial ClpS binds to the N terminal of ClpA (36) and affects substrate specificity. Interestingly, a homologue is also present in the A. thaliana genome (assigned here ClpT), and sequence analysis suggests chloroplast localization (see Table I). The bacterial ClpS protein is not related to ClpS1, which we identified as part of the chloroplast ClpPR complex (3).

The Clp gene family can be divided into three groups of ClpP/R (proteases), ClpC/D/B/F/S/Z/X (chaperones), and ClpT (modulator). The chaperone genes are predominantly located on chromosome 5, whereas the ClpP, -R, and -T genes are mostly on chromosome 1 (Fig. 1A). Analysis of the mitochondrial and plastid chaperones (ClpC1,2/D/B3/S1,2/1,2,3) shows conservation of a I/L/GF motif combined with an upstream basic residue (Lys/Arg), in all sequences but ClpB3 (Fig. 1B). In
greater stability of several chloroplast protein complexes. All stained protein spots were analyzed by a combination of matrix-assisted laser desorption ionization time-of-flight MS (MALDI-TOF MS) and nano-LC-ESI-MS/MS. The purification of Clp complexes, their gel-based separation, and MS analysis were repeated several times. The relevant MS/MS data are listed in Table 1, A and B, of the Supplemental Material, whereas complementary MALDI-TOF MS data are not shown. The criteria for positive identification are described under "Experimental Procedures."

We have indicated the Clp complex in Fig. 2, A and B, and also numbered the most abundant proteins around the Clp complex. About 95% of the Clp core was soluble in the stroma, and 5% was associated with the membrane. In both complexes, four ClpR proteins (ClpR1–4), two ClpS proteins (ClpS1,2), as well as five of six ClpP proteins (ClpP1 and ClpP3–P6) were identified by MS. Thus we now also identified ClpS2 and ClpR4 in the core complex. ClpP2 was also identified in the thylakoid-associated complex but not in the stromal complex. Further below, we demonstrate that this was because of a minor mitochondrial contamination. Protein spots in the area around the Clp complex were all identified by MS and were unrelated to the Clp machinery (Fig. 2, A and B).

The ClpC chaperone was identified in the stroma as a dimeric complex of ∼200 kDa (Fig. 2C). A very large number of sequence tags generated by MS/MS were identified, matching to both the C1 and C2 protein (Table 1C, Supplemental Material). A number of spots were unique for ClpC1, and none were unique to ClpC2. Thus the dimer certainly contained ClpC1 and probably not (or very little) ClpC2. We searched for ClpD and ClpB3, and a small gel spot with ClpB3 was found in the interface of the stacking gel of the CN-PAGE dimension (above 1000 kDa) (not shown).

Separation of Stromal Clp Complexes by Sucrose Gradient Fractionation and Native Isoelectric Focusing (NIEF)—The chloroplast Clp complex always migrated on both the BN-PAGE and CN-PAGE gels as a single heteromeric complex of ∼325 kDa. However, it is still possible that the complex observed in the gels is in fact a mixture of different complexes with approximately the same mass. To elucidate this important point, stromal complexes were separated based on their native pl values, rather than native mass. If indeed Clp core complexes of different composition exist, they are likely to migrate at different pl values, because the predicted pl of the different processed ClpP,R and S subunits vary widely between 4.7 and 8.6 (Table II).

Stromal complexes were first separated by sucrose density gradient centrifugation, to allow removal of the majority of the very abundant 550-kDa ribulose-bisphosphate carboxylase/oxygenase complex. This helped to improve the identification of a possible heterogeneous Clp core population. The fractions containing the Clp complexes (between ∼250 and 450 kDa) were collected and separated on native IEF tube gels with a pl from 3 to 10, followed by SDS-PAGE. Extensive MALDI-TOF MS and nano-LC-ESI-MS/MS analysis of all visible protein spots indicated that all Clp complexes migrated between pl 4 and 6 in a narrow pl range (data not shown). The sucrose gradient fractions were then analyzed on a narrower pl range gel (from 3.5 to 7) to improve resolution. Fig. 3A shows the chloroplast Clp complex migrating at pl of ∼5.0. Extensive MS analysis showed that it contained the same set of 11 Clp proteins, similarly as the ∼325-kDa complex separated based on native mass (Fig. 2A). This suggests that there is a single Clp core complex with the same subunit composition. It is, however,
contribute to staining intensity using silver nitrate or Coomassie.

with the N-terminal domain of eClpA (1K6K), as well as predicted pI and the number of residues in the mature forms of the Clp proteins, that have been separated, based on either native mass or native pI value. But it is unlikely that there are Clp complexes with a composition that would give rise to a strong pI shift, because they would likely have minor differences in composition that do not affect the native pI value. However, ClpR2 has a total of 42 Lys, Arg, His, and Tyr residues, whereas ClpP4 has only 12. This suggests that possible that there is a mixture of Clp core complexes with the same composition but with different spatial arrangements (positions of the Clp subunits in the rings) or, less likely, that minor differences in composition do not affect the native pI value. But it is unlikely that there are Clp complexes with a strongly reduced complexity, because they would likely have been separated, based on either native mass or native pI value.

**Stoichiometry of Plastid ClpP/R/S Proteins—**The E. coli eClpP core is formed by two rings of seven eClpP proteins each. Based on our modeling work in this paper (see further below), we infer that the plastid ClpP/R proteins form a similar ring structure but instead are peripherally associated. Thus the plastid ClpP/R core must contain 14 ClpP/R proteins. Given the presence of 9 different ClpP/R proteins, this means that a number of ClpP/R proteins are present in more than one copy. We thus set out to determine the stoichiometry of the core subunits in the Clp complex. This is a rather challenging task because the 11 proteins migrate very closely together in the range of 22 to 29 kDa. Stromal Clp cores were separated by CN-PAGE, followed by high resolution SDS-PAGE with 9–15% acrylamide (Fig. 3B). Coomassie staining and nano-ESI-MS/MS analysis indicated that ClpS1, ClpS2, and ClpR2 now separated as individual spots. The integrated staining intensities (staining volumes) of these pure proteins were quantified, and spots were subsequently identified by mass (38). Both images were quantified, and spots were subsequently identified by nano-ESI-MS/MS. Now ClpP2, ClpR4, and ClpP4 were separated as individual spots, whereas ClpP1, -P5, and -P6 were still partially overlapping, as were ClpR3, -R1, and -R3. The staining volume of ClpR2 was about 30% of ClpR4 and -P4. The staining volume of ClpR4 and ClpP4 and their molecular masses are approximately identical. However, ClpR4 has 21 lysine residues, whereas ClpP4 has only 12. This suggests that the molar ratio between ClpP4 and ClpR4 is about 2:1. ClpR2 has 11 Lys residues, nearly the same as ClpP4, and ClpR2 is only 1–2 kDa smaller than ClpP4. Given the 30% staining volume of ClpR2 as compared with ClpP4, this suggests that the molar ratio between ClpP4 and ClpR2 is about 3:1. In our earlier paper (3), we suggested that ClpP5 is probably more abundant than ClpP6 and ClpP1, which was based on multiple strongly dominating peptide ions derived from ClpP5 in both MALDI-TOP MS and ESI-MS/MS. Indeed as shown in Fig. 3C, ClpP5 was found in three adjacent spots, whereas P1 and P6 were found together in one spot. Most likely, ClpP1 and -P6 run closely together in this spot, with the presence of ClpP5 being a minor contamination from the adjacent pure ClpP5 spots. Considering that ClpP5 has only 8 Lys and that the total staining volume of the two spots with pure ClpP5 is about 50% of those of R4 and P4, we suggest that the molar ratio between ClpP5 and ClpP4 is at least \( 0.5 \times 12.8 \times 2723 = -0.881 \). Taken these data together, we conclude that the approximate molar stoichiometries between P4:P5:R4:S1:S2 is about 3:3:2:1:1:1.

Considering all gels, the low lysine content of ClpP1 and

### Table II

| protein | pI | C-terminal extension | (Lys + Arg + His + Tyr) | Sequence identity | Aligned/gap positions | r.m.s.d. | C° |
|---------|----|----------------------|--------------------------|-------------------|-----------------------|---------|----|
| ClpP1  | 4.98 | +12 | 6.27 | 43.6 | 181/1 | 0.4 |
| ClpP2  | 5.88 | +21 | 10.35 | 53.8 | 173/0 | 0.5 |
| ClpP3  | 5.24 | +52 | 17.34 | 53.8 | 173/0 | 0.6 |
| ClpP4  | 4.68 | +41 | 12.30 | 49.7 | 173/0 | 0.5 |
| ClpP5  | 5.46 | +12 | 8.32 | 61.3 | 173/0 | 0.5 |
| ClpP6  | 7.9 | +4 | 10.31 | 41.9 | 172/0 | 0.5 |
| ClpR1  | 8.62 | +33 | 21.64 | 39.1 | 174/10 | 0.7 |
| ClpR2  | 8.59 | +16 | 11.42 | 38.5 | 182/0 | 0.4 |
| ClpR3  | 5.89 | +23 | 16.45 | 35.2 | 176/11 | 0.8 |
| ClpR4  | 8.44 | +17 | 21.45 | 45.5 | 176/8 | 0.9 |
| ClpS1  | 5.59 | 18:31 | 18.1 | 138/6 | 0.9 |
| ClpS2  | 6.9 | 16:31 | 21.7 | 138/6 | 1.0 |

\(^a\) C-terminal extension of core subunits as compared with eClpP.

\(^b\) Root mean square deviations computed using the positions of the C\(^{\text{r}}\) atoms after optimal superposition of these atoms on the target and template molecules.

\(^c\) The predicted mass of the mature protein (38.2 kDa) is much larger than the apparent mass (28 kDa).

\(^d\) Structural neighbor of eClpP.
Clp Complexes in Plastids and Mitochondria

Fig. 3. Separation of stromal Clp components by IEF and their stoichiometry. Proteins were identified by MS. A, chloroplast stroma was fractionated on sucrose gradients, and the fraction from 250–450 kDa was separated by NIEF. After focusing, protein complexes were run on SDS-PAGE, and proteins were visualized by MS-compatible silver staining. ESI-MS/MS data for Clp spots 1–6 are summarized in Supplemental Material Table 1. B, chloroplast stromal complexes were separated by CN-PAGE, followed by denaturation and separation on SDS-PAGE with 13–20% acrylamide and stained with Coomassie. C, the Clp core was separated first on SDS-PAGE from 9 to 15% acrylamide, then incubated with SDS and urea, and loaded on a 13–20% acrylamide SDS-PAGE and stained with silver nitrate.

ClpP6, and the comparable number of sequence tags obtained by MS/MS from both P1 and P6, we speculate that ClpP1 and -P6 are present at approximately equal molar ratios to each other and ClpS1 and ClpS2. ClpP3,R1,R3 clustered together in most gels and the staining volume both in Coomassie and silver stain seems quite similar or somewhat less than of ClpP4 + -R4. We tentatively suggest that the molar ratio between the sum of P3 + R1 + R3 is comparable with the sum of ClpP4 + -R4. This roughly leads to a approximate stoichiometry of P4: P5+R4:R2:P3 + R1 + R3:P1:P6:ClpS1:S2 = 3:3:2:1:(3):1:1:1, totaling 16 proteins per core (including S1 and S2).

Clp Cores in Plastids of Roots and Petals in the Genus Brassica—Plastids have a large number of biochemical functions and are present in both photosynthetic tissues (leaves and stems) and non-photosynthetic tissues, such as roots and flowers. The proteome composition between non-photosynthetic and photosynthetic plastids is very different and might require different proteolytic systems. One could expect that the expressed Clp machinery is different in these non-photosynthetic plastid types, as compared with the chloroplast Clp complex as an adjustment to the substrate requirements. We therefore determined the Clp composition in non-photosynthetic plastids from roots and (white) flower petals from B. rapa and B. oleracea, respectively (Fig. 4, A and B). Due to the much larger size of their roots and flowers, respectively, they are a much better source for purification of non-photosynthetic plastids than A. thaliana. These Brassica species and A. thaliana belong to the Brassicaceae family, and their proteomes share a very high degree of protein sequence identity. Alignment of the genomic sequences of several Clp genes in B. oleracea (www.tigr.org/db/e2k1/bog1/) and A. thaliana indicates an average sequence identity of more than 90% between processed Clp orthologues (not shown). We are therefore confident that the correct ("true") orthologue was identified when searching the A. thaliana data base with MS/MS data from B. rapa and B. oleracea proteins.

Plastids were isolated from B. rapa roots and B. oleracea petals (petals were collected manually from plants in the greenhouse) using newly developed protocols. The soluble stromal proteomes from both plastid types were separated by CN-PAGE in the 1st dimension, followed by SDS-PAGE. The ~250 most abundant proteins were identified by MALDI-TOF MS and nano-LC-ESI-MS/MS (not shown). The protein patterns of these two-dimensional gels showed that the Clp protease is the most abundant stromal protease in non-photosynthetic plastids, similarly as in the chloroplast (Figs. 2 and 4).

Fig. 4, A and B, shows the region of complexes between 200–500 kDa, and at this mass resolution, one Clp core was found for both organs. ClpP1–3–6, ClpR1–4, and ClpS1,2 were all identified in these complexes by MS/MS. The MS/MS data are summarized in the Table 2, A and B, of the Supplemental Material. ClpP2 was found in root plastids but not in flower plastids, most likely due to a minor contamination of mitochondria (see below). Similarly as for the chloroplast stroma, ClpC was identified as an ~200-kDa complex in both plastid types (spot 8 in lower panels of Fig. 4, A and B), and similarly as in chloroplast, the ClpC1 (rather than ClpC2) was the main species. ClpB3 was identified as a soluble ~200-kDa dimer in petal plastids (spot 7 in Fig. 4A, lower panel). We do want to point out that glucan phosphorylase (At3g29330) was present as a 320-kDa heat shock proteins and ATP synthase subunits, all unrelated to the Clp machinery (Supplemental Material Table 3).

We then separated stromal proteomes of both non-photosynthetic plastids by NIEF (Fig. 4C). In both plastids, the Clp core complex migrated as a single complex in a very focused manner, strongly supporting the results from CN-PAGE.

The Clp Core in Mitochondria of Potato Tubers—to investigate the Clp machinery in plant mitochondria, potato tubers were used as a rich and established source of pure mitochondria. The soluble matrix proteome was separated by CN-PAGE, and all stained protein spots were analyzed by MS (not shown). ClpP2 was identified by nano-ESI-MS/MS only in one location on the two-dimensional gel (Fig. 5), suggesting that ClpP2 forms an ~320-kDa complex (Supplemental Material Table 3).

The sequence alignment between ClpP2 of A. thaliana and Solanum tuberosum (constructed from overlapping ESTs; e.g. gi 21917555) indicated an 86% identity over 196 amino acid residues, with much less sequence identity to the other A. thaliana Clp proteins (not shown). Protein spots in the area above and below ClpP2 were all identified by MS/MS and were 23/26-kDa heat shock proteins and ATP synthase subunits, all unrelated to the Clp machinery (Supplemental Material Table 3).

2 J.-B. Peltier, G. Friso, A. Rudella, Y. Cai, L. Giacomelli, and K. J. van Wijk, unpublished data.
3). No evidence was found for accumulation of any other Clp-P,R,S proteins, other than ClpP2, as expected from localization prediction (Table I) and Western blotting for ClpP2 (39). Thus, the plant mitochondrial Clp core is homomeric in composition, similar as in other eukaryotes (40). The presence of ClpX chaperones subunits (ClpX1–3) is under investigation.

Because ClpP2 was found in the thylakoid fraction and in root plastids, and because the subcellular localization prediction is ambiguous, we verified localization of ClpP2 by using an approach not dependent on purification of organelles. We therefore cloned the full-length cDNA directly upstream of the gene encoding for green fluorescent protein, and we transformed tobacco leaves with this construct under the control of a strong constitutive promoter by biolistic bombardment. The ClpP2:GFP fusion protein was transiently expressed in tobacco leaves. Cells expressing ClpP2:GFP showed green fluorescence in globular structures identified as mitochondria (Fig. 5B).

Clearly, all the green fluorescence was excluded from the chloroplasts, identifiable by the red autofluorescence. As a positive control for mitochondrial targeted GFP, we used the coxIV: mGFP4 fusion protein from Ref. 41 (Fig. 5C).

Compatibility between the Clp/P/R Proteins and Unique Contributions—To investigate the structural compatibility of the different Clp/P/R proteins, unique structural contributions of the different Clp/P/R/S subunits, and to determine the location of ClpS in the Clp complex, we used the structural information of the homotetradecameric ClpP core complex at 2.3-Å resolution (PDB code 1TYF) (15, 42) and the ClpA N-domain (PDB code 1K6K) at 1.8-Å resolution (43) in E. coli by x-ray crystallography.

The six ClpPs and four ClpRs proteins from A. thaliana are homologous to eClpP for most of the sequence lengths, with the exception of 9 or 10 residue insertions in ClpR1,3,4 and extended C termini for several ClpP/R proteins (Table II and Fig. 6). The x-ray crystal structure of 1TYF was used as a template to build three-dimensional models for all ClpP/R proteins. The 9/10 residue insertions (assigned L1 insertions) occur at equivalent sequence positions with reference to the template 1TYF (positions 70–80, in Fig. 6). To model these insertions and the extended C termini, we identified structural neighbors of 1TYF using the combinatorial extension method (44). Among these structures are the monomeric units of both, 4-chlorobenzoyl-coenzyme A dehalogenase (PDB code 1NZY) and D3,D2-enoyl-CoA isomerase ECI1 (PDB code 1HNO). These proteins show very little sequence identity with eClpP, despite their very high structural similarities with 1TYF (see Fig. 6). Modeling the non-catalytic monomers ClpR1,3,4 (Fig. 7A for ClpR1) indicates that the insertion loop L1 replaces some of the residues forming the oxyanion hole in the catalytic units and lies over the catalytic clefts between the head domains and the handles, i.e. they fill the volume between strands 4 and 9 (strand numbering follows the convention of Ref. 15) that constitutes the substrate-binding pocket in the catalytic units. Sequence alignments of the monomers of 1TYF, 1NZY, 1HNO based on the structural alignment obtained from the combinatorial extension method and the BLAST alignments of the Clp/P/R se-
sequences with the eClpP sequence are shown in Fig. 6. Information about the secondary structure elements (α-helices and β-strands) is also indicated.

The extended C termini of the ClpP/R proteins could be modeled using 1NZY and 1HNO because they share a well defined pattern at their C-terminal ends formed by two to four α-helical fragments. Modeling suggests that the extended C termini present in ClpP3–5 and ClpR1–4 can form α-helical structures that can fold over the top of the barrel-like core structure. An example for a three-dimensional model is shown for ClpR1 superimposed on eClpP (Fig. 7A).

After constructing reliable three-dimensional models for all ClpP and ClpR monomers, we created models for different hetero-tetradecameric plastid Clp cores, as well as for the homo-tetradecameric ClpP2 core from mitochondria (Fig. 7, B and C). All possible contacts between different types of monomers (each unit is in contact with 4 others) were generated and indicated that the different ClpP and ClpR proteins could be part of the same tetradecameric ring structure, without obvious incompatibilities. The root mean square deviations of the Cα atomic positions (Cα r.m.s.d.) obtained after optimal superposition of the structures of the template 1TYF and the ClpP/R models are low (see Table II), showing that the models are highly reliable. All models were carefully analyzed by visual inspection and validated using the program WHAT_CHECK (34).

Putative Lateral Openings in the Clp Core—Sequence conservation of two Arg residues present in all ClpP/R proteins motivated us to look into their three-dimensional environment in 1TYF and the modeled proteolytic ClpPR cores. Careful analysis of this part of the structures suggests putative lateral openings that connect the interior of the central cavity with the outside (Fig. 8, A and B). Two types of putative exits/entrances can be observed.

(i) The first type (Fig. 8, A and B, pink arrows) is located at the interface of the two rings, and it is formed by residues contributed by four different units, i.e. two from each ring (Fig. 8C), with a total of 7 openings per core. These openings measure ~8 Å in diameter and are remarkably polar. The type of residues lining the opening varies, depending on the ClpP/R monomers that contribute to the site. However, two conserved Arg residues are present in all ClpP/R and contributed to these openings by opposite monomers (A and N in Fig. 8A). One of these Arg residue precedes the catalytic Asp residue (Asp-171 using the 1TYF residue numbering) in all ClpP proteins. We also observed that the insertion loops L1 in ClpR1,3,4 can contribute with additional polar residues to flank the internal lips of the polar opening when the monomers are located in any of the two positions shown in blue in the diagram shown in Fig. 8C. All combinations of monomers maintain the polar character of this putative opening with the only exception of ClpR3. A two-residue insertion (Leu-123 and Met-124) in ClpR3 may block or partially close this opening when ClpR3 occupies the two positions shown in red in the diagram (Fig. 8C). We suggest that these polar openings may provide fast exit routes for the proteolytic products (see "Discussion").

(ii) The second type of putative opening (Fig. 7, A and B, yellow arrows) occurs between the two closest catalytic sites located in neighboring monomers on the same ring, thus with 14 putative openings per core. The walls of these putative openings are formed largely by residues located in helices D and E in the first monomer and residues in strands 6, 8, and 10 in the second monomer. A large number of hydrophobic residues flank the internal walls of these putative channels, and fluctuations of the side chains of these residues can provide an oval-like opening with maximum and minimum radii of about 10 and 7 Å, respectively.

![Fig. 5](http://www.jbc.org/)

**Fig. 5. Identification of the ClpP2 core complex from mitochondria of potato tubers.** Gels were stained with Coomassie and proteins identified by ESI-MS/MS and homology-based searches. ESI-MS/MS data for spots 1–10 are summarized in Supplemental Material Table 2. We report here orthologues in *A. thaliana*. Spots 1 and 2, ATP synthase δ (At5g13450); spots 3 and 4, Hsp23-kDa (At4g25200); spot 6 and 7, Hsp26.5kDa (At1g52560); spot 8, cytochrome b, reductase (At5g20080); spot 9, formate dehydrogenase (At5g14780); spot 10, ATP synthase α (At2g07698) and β subunits (At5g08670). B and C, cells expressing ClpP2:GFP showing that ClpP2 (B) and the positive control coxIV:mGFP4 (C) are exclusively targeted to mitochondria. The figures show a projection of 9 transverse optical sections taken at 1-μm intervals along the optical axis. The pseudo colors applied are green for GFP and red for chlorophylls.
Modeling of ClpS1,2 and Their Interaction with the Plastid ClpPR Core—Searches with BLAST and the 3DPSSM threading tool for ClpS1,2 indicate excellent sequence-to-structure alignments (95% certainty) with the N-terminal domain of eClpA (PDB 1K6K), recently deposited by Guo et al. (43). This N-terminal domain is assigned here as the “N domain.” The alignments (Fig. 9) were used to build three-dimensional models of ClpS1,2.

The models of ClpS1,2 (ClpS1 in Fig. 10) show that most of the hydrophobic residues (in green, Fig. 10A) constitute the central core of the structure and are probably important for structural reasons only. On the other hand, conserved residues Arg-18, Glu-28, Arg-57, and Ser-118 may be involved in functional aspects of ClpS. Our conclusions are based on the following: (a) the accessibility of these residues (surface-exposed residues), and (b) on the x-ray experimental structure of the N-terminal domain of eClpA bound to eClpS (PDB code 1MBU) of Ref. 45 that shows the equivalent Glu-28 and Ser-118 (same residue numbers in monomers A and B of 1MBU) interacting with eClpS. The latter x-ray crystal complex also shows that two loops in eClpA provide most of the surface of interaction with eClpS. The equivalent loops in ClpS involve loop LL1 from Arg-18 to Glu-28 in ClpS1 and loop LL2 from Asp-68 to Leu-80 in ClpS2. Residues in these two loops are shown in magenta and orange, respectively (Fig. 10C). It should be noted that only the first loop, LL1, involving residues Arg-18 to Glu-28, is inside the region recognized as the N domain.

Given that the modeled structures of the ClpS proteins are quite different in shape from the ClpP/R proteins, it is highly unlikely that ClpS1,2 is part of the heptameric rings. In contrast, the surfaces of the Clp core and ClpS1,2 show a number of complementarities in shape and hydrophobicity/polarity. Our models of the proteolytic core show the hydrophobic P1 pockets (15) near the axial entrances as possible regions for binding. As we mentioned above, our models for all the ClpP,R monomers, with the exception of ClpP1 and ClpP6, indicate that the P1 pockets may be occluded by the C termini of the monomers forming the site (Fig. 7, B and C, and Fig. 10). When exposed, the P1 grooves always show very good surface complementarity with the epitope formed by loops LL1 and LL2 in ClpS1,2 (Fig. 10, B and C). In addition, the highly aromatic characters of the P1 pockets on the core complex and the ClpS1,2 epitope (with two Tyr, two Glu, and four additional Pro residues; see alignment in Fig. 9) can provide the elements for a tight binding between the two molecules, whereas some of the conserved charged residues, e.g. Arg-18 and Glu-28 in ClpS1,2 and Glu-64 in the ClpP,R, may provide the steering guidance to orient the two molecules during binding. The (I/L)GF motif, as mentioned earlier, might be involved in a different way of interaction.

DISCUSSION

The Surprising Complexity of Clp Cores in Plastids and Mitochondria—Chloroplasts have an extensive thylakoid membrane system, and their soluble stromal proteome is dominated by the Calvin cycle enzymes, whereas leucoplasts and root plastids are non-photosynthetic, have no thylakoid membranes, and contain very little Calvin cycle enzymes. Thus these are very different plastid types, with potentially different needs for proteolysis. Because the Clp family in Arabidopsis and other higher plants is much more complex than in bacteria, one can postulate that the Clp genes are differentially expressed in dependence of tissue type. In this study, we have therefore analyzed the Clp machinery in these three very different plastid types, as well as in mitochondria.

Surprisingly, in all three plastid types, we always detected a Clp complex of ~325 kDa containing 11 different Clp proteins,
namely ClpP1, ClpP3–6, ClpR1–4, and ClpS1–2. ClpR4 and ClpS2 had not yet been identified in any Clp complex before. This complex migrated at a pI of 5.0 in native IEF tube gels. Considering all gels and image analysis, amino acid content, and the number of peptide sequence tags from MS/MS, we suggest a tentative stoichiometry of P4:P5:R4:R2:(P3/R1/R3):P1:P6:S1:S2/3:3:2:1:(3):1:1:1:1, totaling 16 proteins per core (including S1 and S2). The take-home message is that the analysis clearly excludes that plastids contain significant levels of homomeric cores or cores with only a small number of ClpP/R forms, because they would separate based on either their native mass or pI.

In contrast, the mitochondrial Clp complex is a 320-kDa homotetradecamer of ClpP2, without other ClpP/R orthologues and without association of ClpS1,2. Mitochondrial localization was independently confirmed by transient expression essays of a ClpP2-GFP fusion in leaves, in agreement with Western blot data (39).

Thus all plastid-localized ClpP/R/S gene family members are expressed in roots, leaves, and petal plastids and are assembled into heteromeric structures. The presence of all plastid-localized ClpP/R proteins in one well focused complex, as shown by two different separation principles (by native mass and native IEF), is surprising and suggests that each paralogue is required for structure and/or function. This suggestion is supported by preliminary data from T-DNA insertion lines of *A. thaliana* indicating that several of the nuclear-encoded ClpP and ClpR proteins are essential. The chloroplast-encoded ClpP1 is also known to be essential in tobacco (4, 5) and *C. reinhardtii* (6). This also implies that regulation of plastid and mitochondrial proteolysis is not through regulation of Clp gene expression but rather through substrate recognition mechanisms and interaction of chaperone-like molecules (ClpS1,2 and others) to the Clp core. Western blot analysis of protein extracts of different plant tissues indicated constitutive expression of the ClpP proteins (8, 10), without significant expression changes under different stress conditions. This is in agreement with our results of a single ClpP/R complex and suggests that the heteromeric ClpP/R cores observed in our study do not change their composition significantly when plants are exposed to stress.

Thus to understand regulation of Clp proteolysis, it will be critical to obtain detailed insight in the ClpP/R core structure, and its association with ClpS1,2 and other chaperones. Therefore, homology models were created for all ClpP and ClpR monomers and organized in different configurations in tetradecameric complexes. This showed that the different ClpP and ClpR proteins could be part of the same heptameric ring structure and that there were no strong incompatibilities between heptameric rings of different composition forming tetradecameric complexes.

The models were further used to determine how the different ClpP/R proteins contribute to the structures and potentially affect proteolysis, as well as how they affect interactions with the chaperones and ClpS1,2. From a structural perspective, the obvious difference between the ClpP proteins is the length of their C termini. In particular, the C termini of ClpP3 and ClpP4 are 52 and 41 amino acid residues longer than ClpP1.
and eClpP. Modeling based on the structural neighbors of 1TYF suggests that these C termini may form \( \alpha \)-helical structures that can fold over the axial side of the core into a specific P1 pocket (or new tripeptide motif), thereby potentially interfering with docking of the hexameric chaperones. This is a novel feature for the plastid ClpPR core. There are no other differences between the ClpP isoforms that can be directly related to function or structure, but it is important to point out that the overall sequence identities between the ClpP proteins are not more than 60%.

The presence of the four non-catalytic ClpR proteins is very intriguing. What is the difference between the four gene products and why are these non-catalytic subunits part of the complex? Clearly, the ClpR proteins have lost the catalytic protease site, and they do therefore not directly contribute to proteolysis themselves. However, three of the four ClpR proteins (ClpR1,3,4) have a 9 or 10 amino acid insert (the L1 inserts), possibly affecting the presentation of the substrate to the neighboring catalytic triads in the ClpP proteins. In addition, the L1 inserts do contribute residues to the putative lateral polar openings that might stimulate (ClpR1,4) or inhibit (ClpR3) exit of peptide fragments from the core (see more below). The four ClpR proteins differ also in the length of their C termini, similarly as the ClpP subfamily and can potentially influence chaperone interaction to the ClpP/R core. Thus the ClpR proteins likely have a regulatory role. It is important to note that the actual mechanism of proteolysis is not known for any type Clp core complex in any organism, and it is therefore difficult to predict the impact of the presence of the ClpR proteins on efficiency of proteolysis, as well as the size of the proteolytic end products. It is conceivable that some of the differences in the primary sequence of the ClpP/R proteins are needed for the assembly process of the core, especially because there are no strong structural constraints in varying the ClpP/R composition.

The two axial pores of the Clp proteolytic core have been suggested as entrances to the central chamber, as well as exits for proteolytic products (15). This axial exit route requires that
either the chaperones are removed from the core prior to product release or that one side of the core remains unoccupied. Conservation of two Arg residues in all ClpP/R and eClpP and re-analysis of the 1TWF and the three-dimensional models of plastid and mitochondrial cores suggested conserved putative highly polar lateral openings formed by residues belonging to four subunits located at the interface between the two rings. These polar openings seem to be better candidates as a route to clear the products of catalysis, due to their proximity to the catalytic sites, thus avoiding two-way traffic through the axial pores. All combinations of ClpP/R isomers (with the exception of ClpB3) maintain the polar characteristics of this opening. A second type of putative openings is less attractive and may be less likely to be real openings, given the hydrophobicity of the residues lining the putative openings.

We showed that Clp cores in all plastid types (but not in mitochondria) contain ClpS1,2; the stoichiometry analysis suggested (on average) one copy of each ClpS per core. This indicates that the interaction between ClpS and the core is quite strong, because it “survives” both native PAGE and native IEF. The three-dimensional models strongly indicate that ClpS1,2 do not fit in the tetradecameric ring structure but that they do fit well on the axial sites of the ClpPR cores. This interaction involves the hydrophobic (aromatic) P1 pocket in the core and the S1,2 loops LL1 and LL2 and a number of additional conserved residues in ClpS1,2. This predicted interaction site does not fit well on the axial sites of the ClpPR cores. Interestingly, we did clearly identify ClpC1 but not ClpC2. ClpC1 was also found to associate transiently with the inner envelope when involved in proteolysis (16), and chloroplasts (for ClpC) (46). We found that the concentration of ClpC1 was in the same order of magnitude as the concentration of the ClpP/R core. Interestingly, we did clearly identify ClpB1 but not ClpB2. ClpB1 was also found to associate transiently with the inner envelope when involved in protein import (47, 48), although the majority is located in the soluble stroma, as determined by Western blots (8). We are not aware of any study in which the ClpB2 protein was identified specifically. The presence of this second ClpC gene only became clear recently, after sequencing of the A. thaliana genome. We did not find expression of the chaperone ClpD (also named ERD1), which is not surprising because it is reported to be expressed during drought stress and/or during senescence, although the literature is not consistent (compare Refs. 9 and 49). ClpB3 was identified as a 200-kDa dimer in petal plastids, and it overlapped with a 200-kDa dimer of glucan phosphorylase, visible in all three plastid types. This might have masked ClpB3 in chloroplasts and root plastids. However, based on the data from the E. coli homologue and the absence of the “IGF” motif, ClpB3 is unlikely to interact with the Clp core and is not playing a role in Clp-mediated proteolysis.

Conclusions for Clp Proteolysis in Higher Plant Plastids and Mitochondria—The Clp machinery in plastids of higher plants has acquired a surprising complexity. In contrast, the much simpler Clp machinery in plant mitochondria still resembles an E. coli eClpP core. Together, our data show that Clp core complexes present in different plastid types have a similar size and overall composition and that the Clp complex is the most abundant protease complex in the stroma. The essentiality of ClpP1 in tobacco and C. reinhardtii (4, 6), and other ClpP and ClpR genes in A. thaliana, together with the Clp abundance and complexity, suggests a very central role for Clp in plastid homeostasis, equivalent to the central role of the proteasome in the cytosol.

Our findings also show that regulation of plastid proteolysis by the Clp machinery is not through differential regulation of ClpP/R/S gene expression. Instead proteolysis could be regulated by substrate recognition mechanisms and interaction of chaperone-like molecules (ClpS1,2 and others) to the ClpP/R core. This is why the detailed insight obtained by three-dimen-

---

8 O. Vallon, personal communication.
Fig. 10. ClpS1,2 structure and modeled Clp core interactions. A, stereo view of a model of ClpS1. The model was generated by using the structure of the N-terminal domain of eClpA as the template and the sequence alignment obtained using the three-dimensional PSSM threading tool. Only those residue side chains satisfying conservation of hydrophobic, charged, or polar properties are displayed. Color code see the legend to Fig. 9. Conserved residues are in thicker lines. B, hypothetical mode of interaction of ClpS with the Clp core. A solvent-accessible surface representation of a ClpS1 monomer (in magenta) bound into one of the P1 pockets formed at surface of two ClpP,R monomers (ClpP3 in red and ClpP6 in orange) in the catalytic core. The remaining 12 ClpP,R monomers in the Clp core are represented using green ribbons. C, another stereo view of the “docking” model of ClpS1 into the P1 pocket formed by the two neighbor ClpP3 and ClpP6 monomers using an atomic representation of residues that may be involved in the interaction. The remaining 12 ClpP,R monomers forming the catalytic core were omitted. Asp-68 indicates the position of a conserved Asp residue in the ClpRs. Similarly, Arg-18 corresponds to a conserved residue in ClpS. The N terminus of the monomer is indicated by N. The model of ClpS1 bound to the Clp core was generated manually by rigid docking of the backbones of both structures, followed by subsequent modification of the positions of few residue side chains to overcome atomic clashes.

Functional homology modeling of the Clp machinery is so important, as it points to several regulatory mechanisms. Given the essential nature of most (if not all) ClpP/R genes, the assembly of all ClpP/R subunits in one complex and the expression of the ClpP/R machinery in both photosynthetic and non-photosynthetic plastids, it will require a very significant effort to identify Clp substrates, mechanisms of substrate recognition and proteolysis, and the functional role of contribution of all Clp family members. Clearly, the combined experimental and theoretical analysis and the resulting working models for the Clp machinery presented in this paper will form a strong basis for future experimental studies.

Acknowledgments—Dr. David Christopher is gratefully acknowledged for the advice on purification of A. thaliana chloroplasts. We thank Dr. Mike Nasrallah for growth of B. oleracea, Dr. Vladimir Zahroukov for help with the purification of mitochondria, and Heidi Rutschow for sucrose gradient fractionation. Members of the Dr. Maureen Hanson laboratory (Garret Anderson) are gratefully acknowledged for vectors, tobacco plants, and advice on the GFP experiments. Drs. Andre Jagendorf and Sondra Lazarowich are acknowledged for native IEF equipment. We thank Drs. Oliver Vallon and Wojciech Majoran for stimulating discussions concerning the Clp machinery. Financial support for instrumentation and software was provided by National Science Foundation Grant MCB 0009042 and NYSTAR (to K. J. v. W.). All modeling research was conducted using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners.

REFERENCES
1. 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
2. Adam, Z., Adamska, I., Nakabayashi, K., Osterretter, O., Haussubl, K., Manuel, A., Zheng, B., Vallon, O., Rodermel, S. R., Shinozaki, K., and Clarke, A. K. (2001) Plant Physiol. 125, 1912–1918
3. Peltier, J. B., Ytterberg, J., Liberles, D. A., Roepstorff, P., and van Wijk, K. J. (2001) J. Biol. Chem. 276, 16318–16327
4. Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T., and Hashimoto, T. (2001) Plant Cell Physiol. 42, 264–273
5. Kuroda, H., and Maliga, P. (2003) Nature 425, 86–89
6. Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M., and Liu, X.-Q. (1994) Mol. Gen. Genet. 244, 151–159
Sokolenko, A., Pojidaeva, E., Zinchenko, V., Panichkin, V., Glaser, V. M., (2000) *Plant Cell* 12, 137–150
8. Shanklin, J., DeWitt, N. D., and Planagan, J. M. (1995) *Plant Cell* 7, 1713–1722
9. Nakabayashi, K., Ito, M., Kiyosue, T., Shinozaki, K., and Watanabe, A. (1999) *Plant Cell Physiol.* 40, 504–514
10. Zheng, B., Halperin, T., Hruskova-Heidingsfeldova, O., Adam, Z., and Clarke, A. K. (2002) *Physiol. Plant.* 114, 92–101
11. Schelin, J., Lindmark, F., and Clarke, A. K. (2002) *Microbiology* 148, 2255–2265
12. Sokolenko, A., Pojidaeva, E., Zinchenko, V., Panichkin, V., Glaser, V. M., Herrmann, R. G., and Shestakov, S. V. (2002) *Curr. Genet.* 41, 291–310
13. Katayama-Fujimura, Y., Gottesman, S., and Maurizi, M. R. (1987) *J. Biol. Chem.* 262, 4477–4485
14. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C., and Maurizi, M. R. (1998) *J. Biol. Chem.* 273, 12476–12481
15. Wang, J., Hartling, J. A., and Planagan, J. M. (1997) *Cell* 91, 447–456
16. Maurizi, M. R., Thompson, M. W., Singh, S. K., and Kim, S. H. (1994) *Methods Enzymol.* 244, 314–331
17. Burton, B. M., Williams, T. L., and Baker, T. A. (2001) *Mol. Cell* 8, 449–454
18. Missiaen, D., Schwager, F., Betton, J. M., Georgopoulos, C., and Rana, S. (1996) *EMBO J.* 15, 6889–6909
19. Herman, C., Thevenet, D., Boulou, P., Walker, G. C., and D’Ari, R. (1998) *Genes Dev.* 12, 1348–1355
20. Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. (1998) *Genes Dev.* 12, 1338–1347
21. Hoskins, J. R., Yanagihara, K., Mizuuchi, K., and Wickner, S. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 11037–11042
22. Wickner, S., and Maurizi, M. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 8318–8320
23. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., and Baker, T. A. (2003) *Mol. Cell* 11, 671–683
24. Sanchez, R., and Sali, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 13597–13602
25. Baker, D., and Sali, A. (2001) *Science* 294, 93–96
26. Pelizer, J. B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D. A., Soderberg, L., Reesporff, P., von Heijne, G., and van Wijk, K. J. (2002) *Plant Cell* 14, 211–236
27. Zubarevich, V., and Knowles, N. R. (2002) *Arch. Biochem. Biophys.* 402, 136–148
28. Schagger, H., and von Jagow, G. (1991) *Anal. Biochem.* 196, 223–231
29. Schagger, H., and von Jagow, G. (1987) *Anal. Biochem.* 166, 368–379
30. O’Farrell, P. H. (1975) *J. Biol. Chem.* 250, 4687–4921
31. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol.* 234, 779–815
32. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) *Proteins* 23, 318–326
33. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) *J. Mol. Biol.* 299, 499–520
34. Hoef, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* 381, 272
35. Vriend, G. (1990) *J. Mol. Graphics* 8, 52–56
36. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal. Chem.* 68, 850–858
37. Halperin, T., Zheng, B., Itzhaki, H., Clarke, A. K., and Adam, Z. (2001) *Plant Mol. Biol.* 45, 461–468
38. Kang, S. G., Ortega, J., Singh, S. K., Wang, N., Huang, N. N., Steven, A. C., and Maurizi, M. R. (2002) *J. Biol. Chem.* 277, 21095–21102
39. Kohler, R. H., Zipfel, W. R., Webb, W. W., and Hanson, M. R. (1997) *Plant J.* 11, 613–621
40. Wang, J., Hartling, J. A., and Planagan, J. M. (1998) *J. Struct. Biol.* 124, 151–163
41. Guo, F., Maurizi, M. R., Esser, L., and Xia, D. (2002) *J. Biol. Chem.* 277, 46743–46752
42. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng.* 11, 739–747
43. Guo, F., Esser, L., Singh, S. K., Maurizi, M. R., and Xia, D. (2002) *J. Biol. Chem.* 277, 46753–46762
44. Halperin, T., Oesterseter, O., and Adam, Z. (2001) *Planta* 213, 614–619
45. Moore, T., and Keegstra, K. (1993) *Plant Mol. Biol.* 21, 525–537
46. Ko, K., Doung, C., and Ko, Z. W. (1994) *Plant Physiol.* 104, 1087–1098
47. Weaver, L. M., Froehlich, J. E., and Amasino, R. M. (1999) *Plant Physiol.* 119, 1209–1216
Clp Protease Complexes from Photosynthetic and Non-photosynthetic Plastids and Mitochondria of Plants, Their Predicted Three-dimensional Structures, and Functional Implications
Jean-Benoît Peltier, Daniel R. Ripoll, Giulia Friso, Andrea Rudella, Yang Cai, Jimmy Ytterberg, Lisa Giacomelli, Jaroslaw Pillardy and Klaas J. van Wijk

J. Biol. Chem. 2004, 279:4768-4781.
doi: 10.1074/jbc.M309212200 originally published online October 30, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M309212200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2003/12/16/M309212200.DC1

This article cites 49 references, 19 of which can be accessed free at
http://www.jbc.org/content/279/6/4768.full.html#ref-list-1