

# Affinity Purification of the *Arabidopsis* 26 S Proteasome Reveals a Diverse Array of Plant Proteolytic Complexes<sup>\*S</sup>

Received for publication, April 21, 2010, and in revised form, May 27, 2010. Published, JBC Papers in Press, June 1, 2010, DOI 10.1074/jbc.M110.136622

Adam J. Book<sup>†1</sup>, Nicholas P. Gladman<sup>‡2</sup>, Sang-Sook Lee<sup>‡</sup>, Mark Scalf<sup>§</sup>, Lloyd M. Smith<sup>§</sup>, and Richard D. Vierstra<sup>‡3</sup>

From the Departments of <sup>†</sup>Genetics and <sup>§</sup>Chemistry, University of Wisconsin, Madison, Wisconsin 53706

Selective proteolysis in plants is largely mediated by the ubiquitin (Ub)/proteasome system in which substrates, marked by the covalent attachment of Ub, are degraded by the 26 S proteasome. The 26 S proteasome is composed of two subparticles, the 20 S core protease (CP) that compartmentalizes the protease active sites and the 19 S regulatory particle that recognizes and translocates appropriate substrates into the CP lumen for breakdown. Here, we describe an affinity method to rapidly purify epitope-tagged 26 S proteasomes intact from *Arabidopsis thaliana*. In-depth mass spectrometric analyses of preparations generated from young seedlings confirmed that the 2.5-MDa CP-regulatory particle complex is actually a heterogeneous set of particles assembled with paralogous pairs for most subunits. A number of these subunits are modified post-translationally by proteolytic processing, acetylation, and/or ubiquitylation. Several proteasome-associated proteins were also identified that likely assist in complex assembly and regulation. In addition, we detected a particle consisting of the CP capped by the single subunit PA200 activator that may be involved in Ub-independent protein breakdown. Taken together, it appears that a diverse and highly dynamic population of proteasomes is assembled in plants, which may expand the target specificity and functions of intracellular proteolysis.

Plants, like other eukaryotes, rely on the selective removal of abnormal/nonfunctional polypeptides and key short lived regulatory proteins to maintain homeostasis and control their physiology, growth, and development. Arguably, the main protease in plants is the 26 S proteasome, a 2.5-MDa complex responsible for the bulk of ubiquitin (Ub)<sup>‡</sup>-mediated proteolysis (for reviews see Refs. 1, 2). Although the accumulated knowl-

edge of the 26 S proteasome is mainly derived from the analysis of yeast and mammalian complexes, emerging studies indicate that a similar complex exists in plants (3, 4). Its intricate architecture is generated by a 28 subunit core protease (CP) capped at both ends by an 18 subunit or more regulatory particle (RP). The CP is a self-compartmentalized multicatalytic protease created by the assembly of four stacked heptameric rings of  $\alpha$  and  $\beta$  subunits in a  $\alpha_{1-7}/\beta_{1-7}/\beta_{1-7}/\alpha_{1-7}$  configuration. A central chamber encloses the active sites for peptidylglutamyl-peptide-hydrolyzing, trypsin-like, and chymotrypsin-like activities provided by the  $\beta 1$ ,  $\beta 2$ , and  $\beta 5$  subunits, respectively. Access to this chamber is guarded by an axial pore created by each  $\alpha$  subunit ring, which employs a sophisticated gating mechanism to restrict access such that only unfolded proteins may enter the catalytic chamber (2, 5).

The RP binds to one or both ends of the CP and sits directly over the  $\alpha$  ring pore. It is composed of two subcomplexes as follows: a base containing six related AAA-ATPases (designated RPT1–6 for regulatory particle triple-A ATPases) and three non-ATPase subunits (designated RPN1, RPN2, and RPN10, for regulatory particle non-triple-A ATPases), and a lid that contains at least 12 additional RPN subunits (RPN1–3 and -5–13) (1–3). The functions of only some of the RP subunits are known. The six RPT subunits form a ring that consumes ATP to facilitate channel opening, and target protein unfolding and translocation into the CP (5, 6). The RPN10 and RPN13 subunits function as the major Ub receptors to identify appropriate substrates bearing poly-Ub chains (7–10). RPN1 has been shown to interact with ubiquitin-binding shuttle proteins and thus may also assist in substrate recognition (11). RPN11 has a deubiquitylating activity that can remove Ub moieties bound to target proteins during their breakdown (12). The possible roles of the remaining subunits include the assembly and maintenance of structural integrity, substrate selection, processing steps to ready a target for breakdown, target import, discharge of amino acid/peptide products from the CP lumen, and Ub recycling (2, 13).

In addition to core subunits, a number of other factors associate stoichiometrically with the 26 S proteasome. Given the complicated design of this multisubunit particle, it is not surprising that some act transiently as chaperone-like factors to promote assembly (14–16). Others appear to aid in polyubiquitylating substrates (e.g. E3s), the delivery of these substrates to the complex, and/or Ub recycling during substrate breakdown (2, 17, 18). In the most extreme case, novel complexes have been

cle; T, transfer; WT, wild type; RT, reverse transcription; Rubisco, ribulose-bisphosphate carboxylase/oxygenase; HA, hemagglutinin.

\* The MS facility was supported in part by National Institutes of Health Grants 1P50HG004952 from NHGRI and P01GM081269 from NIGMS (to M. S. and L. M. S.). This work was supported by United States Department of Energy Basic Energy Sciences Program Grant DE-FG02-88ER13968 and National Science Foundation *Arabidopsis* 2010 Program Grant MCB-0115870 (to R. D. V.).

<sup>S</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables 1–5, Figs. 1 and 2, and Datasets 1 and 2.

<sup>1</sup> Supported by National Institutes of Health Ruth L. Kirschstein Predoctoral Fellowship F31NS054563.

<sup>2</sup> Supported by National Institutes of Health predoctoral fellowship to the University of Wisconsin Genetics Training Program.

<sup>3</sup> To whom correspondence should be addressed: Dept. of Genetics, 425-G Henry Mall, University of Wisconsin, Madison, WI 53706. Tel.: 608-262-8215; Fax: 608-262-2976; E-mail: vierstra@wisc.edu.

<sup>4</sup> The abbreviations used are: Ub, ubiquitin; CID, collision-induced dissociation; CP, core protease; ETD, electron transfer dissociation; ESI, electrospray ionization; HCD, high energy collision dissociation; HPLC, high performance liquid chromatography; MS, mass spectrometry; RP, regulatory parti-

identified in yeast and mammals that consist of the CP capped on one or both ends by the hexameric PA28/11S regulator (19) or the single proteasome activator (PA)-200 protein (Bln10 in yeast) (20–22).

In the past few years, great strides have been made in defining the 26 S proteasome from plants and identifying most of its core subunits (3, 4). Both genomic analyses and biochemical studies of purified preparations have shown that the complex is more heterogeneous than in most other eukaryotes. For example, whereas each core CP and RP subunit is encoded by a single gene in yeast, most *Arabidopsis* and rice subunits are encoded by gene pairs, some of which are sufficiently different in sequence (23–26) and/or genetic impact (13, 27–32) to suggest distinct functions. Unfortunately, all previous biochemical analyses of plant proteasomes used conventional chromatographic approaches to isolate the particle (25, 26, 33). Based on studies with the yeast complex, these purifications may have missed less tightly bound core and accessory components (e.g. RPN13, UPL7 (Hul5 in yeast), UBP6, ECM29, and PA200) (34, 35). At least for the *Arabidopsis* particle, the time-consuming protocol also generated preparations contaminated with breakdown products for individual subunits (e.g. RPN10), thus compromising conclusions drawn from their analysis (26).

To better define the activity and subunit composition of the 26 S proteasomes in plants, we developed an affinity-based strategy to effectively purify the holoenzyme intact from *Arabidopsis*. Mass spectrometric (MS) analysis of the complex isolated from young seedlings allowed us to more accurately catalog the subunit composition, post-translation modifications, and proteasome-interacting proteins. Particles assembled with all but three of the potential core subunit isoforms were detected, indicating that a heterogeneous collection of 26 S proteasomes is assembled *in planta*. In addition, a number of CP and RP subunits were found to undergo one or more post-translational modifications, including partial proteolytic cleavage, acetylation, and ubiquitylation. Finally, we identified several other proteins that associate with the *Arabidopsis* proteasome, including the PBAC2 assembly chaperonin, the associated DSS1/Sem1/RPN15 protein, the deubiquitylating enzyme UBP16, and the alternative activator PA200. Our results indicate that plants assemble a highly diverse population of proteasomes that likely expands the functionality of these proteolytic particles.

## EXPERIMENTAL PROCEDURES

**Plant Materials and Growth Conditions**—The *pag1-1* (SALK\_114864) and *pa200-1–6* (SALK\_103341, SALK\_095870, SALK\_070184, SALK\_035070, SALK\_007570, and SALK\_151948, respectively) T (transfer)-DNA insertion mutants in the *Arabidopsis thaliana* ecotype Col-0 were identified in the Signal T-DNA collection and obtained from the *Arabidopsis* Biological Resource Center (Columbus, OH). The *pag1-1*, *pa200-2*, and *pa200-3* alleles were tracked by genomic PCR using the T-DNA-specific left border primer (Lba1) in combination with the gene-specific primers P11, P13, and P15, respectively (see supplemental Table 1 for all primer sequences used in this study). RT-PCR of the T-DNA insertion lines used the primer pairs P1 + P2 (*pag1-1*), P12 + P13 (*pa200-2*) and

P14 + P15 (*pa200-3*) in combination with total seedling RNA as described previously (18). The *rpn12a-1* mutant was described by Smalle *et al.* (30). The mutants were backcrossed three times to the Col-0 parent and then made homozygous by selfing. Unless otherwise noted, plants were grown under sterile conditions on solid Gamborg's B-5 growth medium containing 2% sucrose and 0.7% agar (Invitrogen).

For MG132 treatments, plants were grown on growth medium agar containing 2% sucrose for 4 days under long day conditions. Seedlings were then transferred to liquid growth medium containing 2% sucrose and 100  $\mu$ M MG132 (*N*-(benzyloxycarbonyl)-Leu-Leu-Leu-al; Enzo Life Sciences, New York) and incubated for 30 h under continuous light with gentle shaking (28).

**Affinity Tagging and Complementation**—The genomic region encompassing the full coding sequence of *PAG1* plus the 2-kbp sequence upstream of the ATG translation start site was PCR-amplified from *A. thaliana* ecotype Col-0 genomic DNA, using the primer pair P9 and P10, and cloned into the pENTR/D-TOPO entry vector (Invitrogen). The sequence-confirmed clone was recombined in-frame into the appropriate destination vectors containing the various C-terminal fusion tags, including FLAG (DYKDDDDK), Myc (EQKLISEEDL), and HA (YPYDVPDYA) (pEarleyGate302, 303, and 301, respectively (36)). These constructions were transformed into *pag1-1/+* heterozygous plants by the *Agrobacterium*-mediated floral dip method. Double heterozygous T<sub>1</sub> progeny were selected by Basta resistance and confirmed by genomic PCR to contain the *pag1-1* allele and the *PAG1* transgenes using primers P1 + P2 (wild type), P3 + Lba1 (*pag1-1* T-DNA), and P3 + P5 (*PAG1* transgenes). Homozygous *pag1-1* T<sub>3</sub> plants rescued with the various tagged *PAG1* transgenes were confirmed by immunoblot analysis with anti-PAG1 antibodies.

**Immunological Methods**—The full-length *PAG1* cDNA was PCR-amplified using primers P5 and P6 and cloned into the pDONR221 entry vector (Invitrogen). The sequence-confirmed cDNA was recombined into pDEST17 to express the protein with an N-terminal His<sub>6</sub> tag in *Escherichia coli* strain BL21. Following a 3-h induction with 0.4 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside, the expressed protein was collected from the insoluble fraction by centrifugation, dissolved in 8 M urea, and bound to nickel-nitrilotriacetic acid agarose (Qiagen, Valencia, CA) followed by elution with 250 mM imidazole. PAG1 protein was further purified by preparative SDS-PAGE; gel pieces containing the 27-kDa protein (visualized by Coomassie Blue staining) were excised, rinsed, and injected directly into rabbits (Harlan BioProducts, Madison, WI).

PA200 antibodies were generated against a partial fragment encompassing residues 840–1140. The corresponding cDNA sequences were obtained by RT-PCR amplification of total seedling RNA using the primer pairs P7 + P8 and subcloned into pET23b (Novagen, San Diego). The protein fragment was expressed in BL21 cells, purified, and injected into rabbits as described for PAG1. Immunoblot analyses were performed according to Smalle *et al.* (30). Antibodies against CSN4 and CSN5, eIF3b and eIF3-e, HSC70, Rubisco large subunit, and nitrilase-1 were provided by Drs. Xing-Wang Deng, Daniel Chamovitz, Charles Guy, Archie Portis, and Mark Estelle,

## Characterization of the 26 S Proteasome from Arabidopsis

respectively. Antibodies against PBA1, RPN1, RPN5, RPN10, RPN12a, RPT2a, and Ub were as described previously (26, 30). Anti-FLAG M2 antibodies were obtained from Sigma.

**Affinity Purification using PAG1-FLAG—PAG1-FLAG** *pag1-1<sup>-/-</sup>* seedlings were grown 10 days under continuous light in liquid growth medium containing 2% sucrose. The seedlings (5 g of fresh weight) were pulverized at liquid nitrogen temperatures using a mortar and pestle and extracted with 7.5 ml of Buffer A (50 mM Tris-HCl (pH 7.5), 25 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM EDTA, and 5% (v/v) glycerol), with 10 mM ATP, 6 μM chymostatin, and 2 mM phenylmethylsulfonyl fluoride added just before use. Extracts were filtered through Miracloth (Calbiochem) and clarified at 25,000 × *g* for 20 min. The supernatant was immediately applied three times to a column containing 50 μl of anti-FLAG M2 affinity beads (Sigma) and then washed three times with 2 ml of Buffer A. Bound protein was eluted by a 20-min rotation at 4 °C with 250 μl of Buffer A containing 500 ng/μl of the FLAG peptide (DYKDDDDK; University of Wisconsin Biotech Center, Madison, WI). The RP subcomplex was selectively released by incubation of bead-bound 26 S proteasomes with Buffer A plus 800 mM NaCl.

Glycerol gradient centrifugation and activity assay were performed as described previously (13). CP peptidase activity in the absence or presence of MG132 was measured using the substrate succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin (BioMol, Plymouth Meeting, PA) (26). Other protease inhibitors were obtained from Sigma. Nondenaturing PAGE followed the protocol of Yang *et al.* (26). For two-dimensional native/SDS-PAGE, the gel lanes from native PAGE were incubated with 1% SDS and 1% 2-mercaptoethanol for 30 min and then subjected to SDS-PAGE in the second dimension.

For the protease inhibitor assays, seedlings were extracted as above with the inclusion of various inhibitors. Concentrations used were as follows: Roche Applied Science complete mini protease inhibitor mixture, 1 tablet/10 ml; aprotinin, 800 nM; antipain, 100 μM; benzamide, 4 mM; chymostatin, 6 μM; E-64, 10 μM; leupeptin, 10 mM; MG132, 50 μM; *N*-ethylmaleimide, 1 mM; pepstatin, 1 μg/ml; and phenylmethylsulfonyl fluoride, 2 mM. Extracts were incubated at 25 °C for 1 h.

**MS Sample Preparation**—Samples purified by the anti-FLAG affinity chromatography were treated with 50 mM dithiothreitol and 55 mM iodoacetamide, digested overnight at 37 °C with 20 ng/μl sequencing grade trypsin (Promega, Madison, WI), or 25 ng/μl lysyl endopeptidase C (LysC; Wako, Richmond, VA) in 100 mM ammonium bicarbonate (pH 8.0), and then vacuum dried. Samples were reconstituted in 100 μl of 5% acetonitrile and 0.1% formic acid, desalted using C18 pipette tips (OMIX tip C18, Varian, Lake Forest, CA), and dried again. The vacuum-dried samples were reconstituted in 10 μl of 5% acetonitrile and 0.1% formic acid prior to liquid chromatographic separation. For affinity purification after MG132 treatments, plants were grown as stated above for 9 days and then treated for 30 h with 100 μM MG132.

**HPLC-ESI-MS/MS HCD and ETD Analysis**—High energy collision dissociation (HCD) and electron transfer dissociation (ETD) MS analyses employed a capillary liquid chromatography-MS/MS system consisting of a high performance liquid chromatograph (HPLC) (Waters NanoAcquity, Milford, MA)

connected to an electrospray ionization (ESI) FT/ion-trap mass spectrometer (LTQ Orbitrap Velos, Thermo Fisher Scientific, San Jose, CA) (37, 38). A fritless 100 × 365-μm fused silica capillary micro-column was prepared by pulling the tip of the capillary to ~1 μm with a P-2000 laser puller (Sutter Instruments, Novato, CA) and packing the capillary with 10 cm of 5-μm diameter C18 beads (Western Analytical Products, Inc., Murrieta, CA). The peptides were loaded over 20 min at a flow rate of 1 μl/min and eluted over 115 min at a flow rate of 200 nl/min with a gradient of 2–45% acetonitrile in 0.1% formic acid. For the trypsin-digested samples, a full mass scan was performed in the FT Orbitrap between 300 and 1500 *m/z* at a resolution of 60,000, followed by 10 MS/MS HCD scans of the 10 highest intensity parent ions at 45% relative collision energy. The HCD scans were also analyzed in the FT Orbitrap detector at a resolution of 7500. Dynamic exclusion was enabled with a repeat count of 2 over a duration of 15 min and excluded for 30 min. For the LysC-digested samples, the same chromatography and full mass scan parameters were used, with ETD analysis of the 10 most intense ions in the ion trap. The ETD activation time was 70 ms, with charge-state screening enabled and the singly and doubly charged ions excluded from analysis.

**HPLC-ESI-MS/MS CID Analyses**—Proteins purified by anti-FLAG affinity chromatography were separated by SDS-PAGE. Following visualization of the lanes with Coomassie Blue, a total of 17 gel bands was excised and subjected to in-gel trypsin digestion. The tryptic peptides extracted from each gel slice were dried, desalted, and reconstituted as above and loaded onto the C18 capillary microcolumn described above. MS analysis employed a capillary liquid chromatography-MS/MS system consisting of an HPLC connected to an ESI ion-trap MS (Surveyor HPLC and LCQ deca XPplus; Thermo Fisher Scientific, San Jose, CA) (39). The peptides were eluted over 150 min at a flow rate of 300 nl/min with a gradient of 5–80% acetonitrile in 0.1% formic acid. A full mass scan was performed between 400 and 2000 *m/z*, followed by three MS/MS collision-induced dissociation (CID) scans of the three highest intensity parent ions at 45% relative collision energy.

**Arabidopsis Proteome Database Searches**—The acquired MS and MS/MS spectra were searched against the *A. thaliana* ecotype Col-0 protein database (IPI Database version 3.61, September 2009) using the SEQUEST program (Thermo Fisher Scientific). Sequences for trypsin, LysC, and human keratin were added to the database to decrease the false-positive discovery rate. Masses for both precursor and fragment ions were treated as mono-isotopic. Oxidized methionine (+16 Da), carbamidomethylated cysteines (+57 Da), acetylation (+42 Da), myristoylation (+210 Da), and phosphorylation (+80 Da) were included as variable modifications. In addition, a mass shift of 114.1 Da for lysine residues was included as a variable modification to identify possible Ub footprints (40, 41). The initial search also allowed for up to two missed trypsin cleavages given that ubiquitylated lysines are often resistant to digestion, with a follow-up analysis that permitted this cleavage (42). The data were filtered using a 1% false discovery rate (43). For the Orbitrap MS/MS datasets, parent ion masses were initially matched with a mass tolerance of 15 ppm for precursor masses and 0.1 Da for HCD fragments; peptides were then filtered using Xcorr



versus charge state (+1 > 1.5, +2 > 2.0, +3 > 2.25, +4 > 2.5, +5 > 2.75, +6 > 3.0), and an Sp score >200. For the ion trap MS/MS dataset, parent ion masses were initially matched with a mass tolerance of 1.4 Da for precursor masses and 1 Da for CID fragments; peptides were then filtered using Xcorr versus charge state (+1 > 1.5, +2 > 2.0, +3 > 2.5). A minimum identification of two unique peptides was required for confident identification of the parent protein.

**Phylogenetic Analyses**—Genes encoding orthologs for various *Arabidopsis* proteasome subunits were identified by BLAST searches of the Phytozome and NCBI databases (blast.ncbi.nlm.nih.gov) using the *Arabidopsis* protein sequence as the query (accession numbers are provided in supplemental Dataset 1). Orthologs were confirmed by reciprocal BLAST searches of the *Arabidopsis* protein database. Protein sequences were then aligned using ClustalW2 (supplemental Dataset 2). Phylogeny was determined by Bayesian estimation using the Mr.Bayes 3.1.2 program (44) with a mixed model. Specific commands for Mr.Bayes were as follows: lset nst = 6 rates = invgamma. Each group was run for 100,000 generations with sampling every 100 cycles for a total of 1000 samples. The sump and sumt commands were used to tabulate posterior probabilities of positive selection for each amino acid site and to build consensus trees. The first 250 cycles were discarded to remove results of estimations obtained before the process reached convergence. Trees were rooted with obvious non-plant orthologs and were viewed and edited using FigTree version 1.3.1.

**Accession Numbers**—Sequence data from this study can be found in supplemental Dataset 2.

## RESULTS

**Identification of a Null Mutant for *Arabidopsis* PAG1**—To enrich for *Arabidopsis* proteasomes containing the common CP subcomplex, we adopted a strategy first successful in yeast that involves replacing one of the 14 CP polypeptides with a tagged version and then exploiting the tag for rapid affinity purification (34, 35). The most attractive candidate in *Arabidopsis* was the  $\alpha 7$  subunit, which is encoded by the single *PAG1* gene (At2g27020) (23). The 249-amino acid PAG1 protein (27 kDa) has high sequence identity to orthologs from other plants (*Oryza sativa* (84%), *Zea mays* (86%), *Glycine max* (87%), *Populus trichocarpa* (91%), *Physcomitrella patens* (82%), and *Vitis vinifera* (90%)), as well as animals (*Homo sapiens* (61%) and *Drosophila melanogaster* (51%)) and yeast (*Saccharomyces cerevisiae* (Pre10, 49%)). As part of the  $\alpha$  ring, PAG1 does not directly participate in proteolysis but, through its N terminus, plays a role in gating the axial pore that guards the CP catalytic chamber (5, 6). The crystal structure of the yeast CP shows that the C terminus of PAG1 is solvent-exposed (45), thus allowing us to append tags to this end potentially without perturbing CP assembly.

As a first step in PAG1 replacement, we identified a null mutation (*pag1-1*) in the *A. thaliana* ecotype Col-0 T-DNA insertion collection that interrupts the 5th intron (Fig. 1A). Homozygous *pag1-1* progeny could not be generated from a self-crossed heterozygous parent (supplemental Table 2), indicating that the  $\alpha 7$  subunit is essential in plants as it is in yeast

(46). We also did not detect aborted embryos in siliques from selfed heterozygous *pag1-1* plants, suggesting that the barrier occurs early in reproduction. This barrier was pinpointed to a defect in male gametogenesis from the analysis of reciprocal crosses between *pag1-1/+* and wild-type flowers. Whereas the *pag1-1* allele was readily transmitted through the egg, no transmission was detected through pollen (supplemental Table 2).

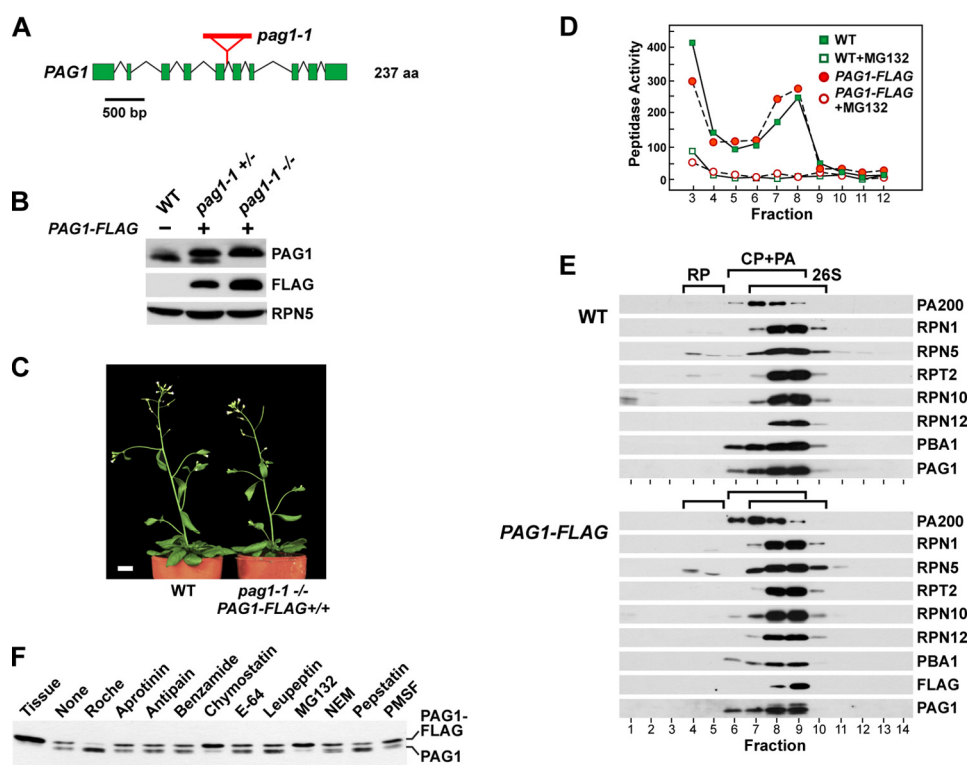
**Complementation of the *pag1-1* Mutant with Epitope-tagged PAG1**—To engineer an affinity-tagged proteasome, we introduced into heterozygous *pag1-1* plants transgenes controlled by the native *PAG1* promoter that express the full-length PAG1 protein bearing a variety of C-terminal epitopes (e.g. FLAG, HA, Myc, and TAP). When we tested for their ability to rescue homozygous *pag1-1* plants in selfed T<sub>2</sub> populations, we failed to identify rescued lines using the *PAG1-TAP* transgene, potentially because this 309-residue tag is too large to retain PAG1 functionality. Fortunately, successful rescue was achieved with the smaller tags, FLAG, HA, and Myc, which added only 23, 23, and 24 residues, respectively. Immunoblot analysis of these lines with anti-PAG1 antibodies confirmed the substitutions. Whereas a doublet of PAG1 protein could be seen in heterozygous *pag1-1* plants expressing the tagged *PAG1* transgenes, only the higher molecular mass, tagged versions could be seen when the *pag1-1* allele was made homozygous (see Fig. 1B for an example). *pag1-1* plants rescued with either *PAG1-FLAG*, *PAG1-HA*, or *PAG1-Myc* were phenotypically indistinguishable from wild type when grown under normal conditions and were fully fertile (Fig. 1C), strongly suggesting that these tagged proteins faithfully replaced their wild-type counterpart.

Initial attempts at affinity purification suggested that the FLAG epitope provided the most robust enrichment and was used in all the subsequent studies. Importantly, the PAG1-FLAG protein successfully integrated into the 26 S proteasome complex without perturbing CP/RP assembly. When crude extracts from wild-type plants were subjected to glycerol gradient centrifugation, we could detect the 26 S proteasome as a peak of peptidase activity using the substrate succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin (26), which was sensitive to the CP inhibitor MG132 (Fig. 1D). This activity co-sedimented with known subunits of the CP and RP as judged by immunoblot analyses with a library of anti-CP (PAG1 ( $\alpha 7$ ) and PBA1 ( $\beta 1$ )) and anti-RP antibodies (RPN1, RPN5, RPT2, RPN10, and RPN12a) (Fig. 1E). Nearly identical sedimentation profiles for peptidase activity and for CP and RP subunits were observed for extracts generated with *PAG1-FLAG pag1-1* plants (Fig. 1E). This co-sedimentation also included the FLAG epitope attached to PAG1-FLAG protein.

During the glycerol gradient fractionations, we noticed that the PAG1-FLAG protein is sensitive to post-homogenization proteolysis, which removes over time most, if not all, of the FLAG sequence. Tests with a variety of protease inhibitors discovered that this *in vitro* cleavage is sensitive to the chymotrypsin inhibitor chymostatin and MG132 but not to many others (Fig. 1F). Consequently, chymostatin was routinely included in the purification protocol to minimize release of the FLAG tag.

**Affinity Purification of the 26 S Proteasome**—Using a homozygous *PAG1-FLAG pag1-1* line, we developed a rapid and robust protocol to enrich for the 26 S proteasome directly

## Characterization of the 26 S Proteasome from Arabidopsis



**FIGURE 1. Arabidopsis PAG1 gene structure and mutant rescue.** *A*, diagram of the single *PAG1* gene encoding the  $\alpha 7$  subunit of the CP and location of the *pag1-1* T-DNA insertion. Boxes indicate exons, and lines indicate introns. aa, amino acids. *B*, expression of the *PAG1-FLAG* protein in homozygous *pag1-1* mutants. Crude extracts from WT plants and heterozygous ( $\pm$ ) and homozygous ( $-/-$ ) *pag1-1* plants expressing the *PAG1-FLAG* transgene were subjected to SDS-PAGE and immunoblot analysis with antibodies against *PAG1*, *FLAG*, and *RPN5a* (loading control). *C*, rescue of homozygous *pag1-1* plants with the *PAG1-FLAG* transgene. Shown are 6-week-old plants homozygous for both the *pag1-1* mutation and the *PAG1-FLAG* transgene grown under a long day photoperiod (16 h light/8 h dark). Bar represents 12 mm. *D* and *E*, assembly of the *PAG1-FLAG* protein into proteasome complexes. Crude extracts from WT and homozygous *PAG1-FLAG pag1-1* plants were fractionated by glycerol gradient centrifugation and then assayed enzymatically for CP peptidase activity (*D*) and for various proteasome subunits by immunoblot analysis (*E*). Sedimentation positions of the 26 S proteasome, free RP, and the PA200-CP (PA-RP) complexes are indicated by the brackets. *F*, efficacy of protease inhibitors in stabilizing the *PAG1-FLAG* protein *in vitro*. Crude extracts from 10-day-old homozygous *PAG1-FLAG pag1-1* seedlings were incubated for 2 h at 25 °C with various protease inhibitors and then analyzed by immunoblot analysis with anti-*PAG1* antibodies. The positions of the *PAG1-FLAG* and *PAG1* minus the *FLAG* tag are indicated. *PMSF*, phenylmethylsulfonyl fluoride.

from crude seedling extracts. It involved an initial homogenization of liquid-grown seedlings in a nondenaturing buffer containing chymostatin, clarification to remove cell debris, and incubation of the soluble material with beads conjugated with anti-FLAG antibodies. Following a wash with a low salt (25 mM NaCl) buffer, proteasomes were competitively released from the antibodies under gentle, nondenaturing conditions simply by incubating the beads with an excess of the 8-amino acid FLAG peptide (DYKDDDDK). Ten mM ATP was routinely included in all buffers to maintain association of the RP and CP subcomplexes (26, 47). Approximately 12  $\mu$ g of proteasomes could be isolated from 5-g fresh weight of seedlings. In total, the procedure took  $\sim$ 2 h, which is substantially faster than the 2-day conventional protocol described by us previously that involved two precipitation steps and two chromatography steps, one of which employed a high salt elution that can release loosely bound subunits/cofactors (26, 35).

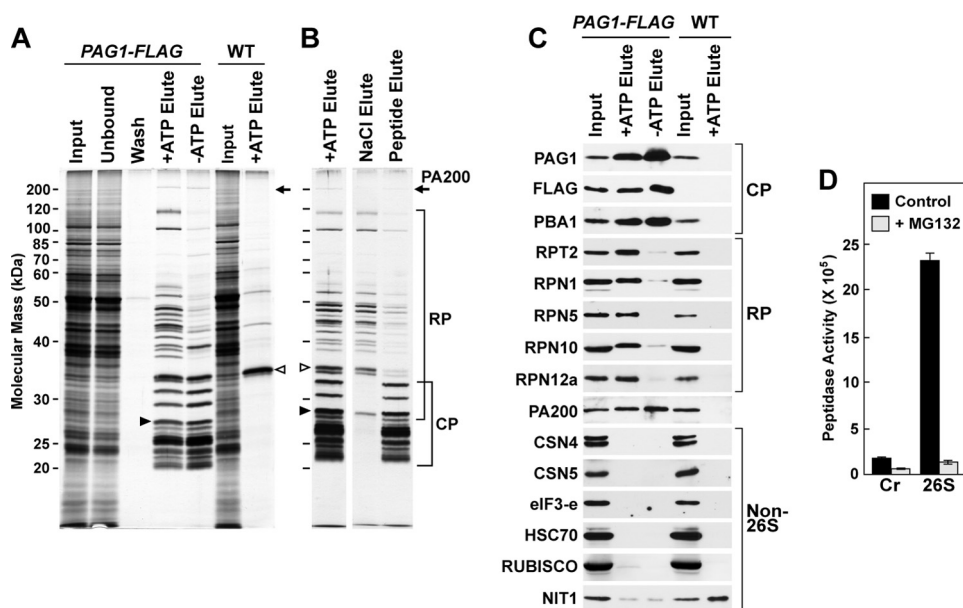
SDS-PAGE and peptidase assays demonstrated that this single-step affinity purification is highly effective in isolating reasonably intact 26 S proteasomes. As compared with a FLAG-peptide eluate generated with wild-type plants, that obtained

from *PAG1-FLAG pag1-1* plants was enriched for MG132-sensitive peptidase activity (Fig. 2D), and for a collection of proteasome proteins reminiscent of that obtained by conventional methods (26), including a 24–32-kDa set that encompasses the  $\alpha$  and  $\beta$  subunits of the CP and a set between 24 and 110 kDa that encompasses most RP subunits (Fig. 2A). Importantly, two abundant contaminants of conventionally purified proteasomes (tripeptidyl peptidase-II and fatty acid  $\alpha$ -dioxxygenase-1 (26, 48)) were absent in the affinity-purified preparations. In agreement with the need for ATP to maintain RP-CP binding and the sensitivity of this association to high salt (34, 35), the abundance of the RP subunits was substantially reduced if ATP was omitted from all steps of the purification (Fig. 2A), or if proteasomes bound to the anti-FLAG antibody resin in the presence of ATP were first washed with high salt (800 mM NaCl) before FLAG peptide elution (Fig. 2B).

Subsequent immunoblot analyses confirmed the presence of specific subunits for both the CP and RP (Fig. 2C). In particular, full-length RPN10 without substantial contamination by breakdown products was detected in the affinity preparations, which contrasted results from the conventional method

where most RPN10 was degraded during purification (26). In agreement with the need for ATP to maintain 26 S proteasome integrity, levels of the RP subunits (RPN1, RPN5, RPN10, RPN12a, and RPT2) but not the CP subunits (*PAG1* ( $\alpha 7$ ) and *PBA1* ( $\beta 1$ )) were substantially lower when we attempted to purify the complex without ATP (Fig. 2C).

In addition to expected CP and RP subunits, we detected several other proteins that could represent: (i) other, yet to be identified, components of the *Arabidopsis* 26 S complex based on studies with yeast particle; (ii) assembly chaperonins; (iii) various accessory proteins; or (iv) other complexes previously shown (eIF3 complex) or proposed (COP9 signalosome (CSN)) to interact with the RP or CP (14, 21, 34, 35, 49–51). The unidentified protein at 200 kDa (Fig. 2, A and B) was of special interest because it matched the molecular mass of PA200, a previously described activator of the CP in mammals (22) and yeast (*Blm10* (21, 52)). Its identity as PA200 was subsequently confirmed using antibodies generated against a portion of the 1781-residue *Arabidopsis* protein (Fig. 2C). The abundance of PA200 in the affinity preparations was unaffected by omitting ATP during the purification, implying that its association with



**FIGURE 2. Affinity purification of proteasomes from *PAG1-FLAG pag1-1* plants.** *A*, SDS-PAGE analysis of the affinity purification steps. Total protein extracted from 10-day-old WT and *PAG1-FLAG pag1-1* plants was incubated with anti-FLAG affinity resin, washed, and competitively eluted with the FLAG peptide. The procedure was performed in the presence or absence of ATP. The input, unbound, washed, and eluted fractions were subjected to SDS-PAGE, and the gel was stained for protein with silver. The *arrow* and *closed arrowhead* locate the PA200 and the PAG1-FLAG proteins, respectively. The *open arrowhead* identifies nitrilase, which is nonspecifically enriched during the purification. *B*, salt dissociation of the 26 S proteasome into the RP and CP subcomplexes. 26 S proteasomes were bound to the anti-FLAG resin in the presence of ATP and either eluted with the FLAG peptide or first eluted with 800 mM NaCl, followed by elution with the FLAG peptide. The *brackets* locate subunits of the CP and RP subcomplexes. The *arrow* indicates PA200. *C*, immunoblot detection of various proteasome subunits in the affinity-purified preparations shown in *A*. Subunits tested include the CP subunits PAG1 and PBA1, and the RP subunits RPT2, RPN1, RPN5, RPN10, and RPN12a. Other proteins tested include PA200, the CSN4 and CSN5 subunits of the CSN complex, the eIF3-e subunit of the eIF3 complex, HSC70, Rubisco small subunit, and nitrilase (NIT1). *D*, peptidase activity of affinity-purified proteasomes. Peptidase activity in the presence or absence of MG132 was measured in the crude extract (*Cr*), and in preparations purified in the presence of ATP, using the substrate succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin. Activities were normalized to total protein concentration.

the CP is ATP-independent (Fig. 2C). Conversely, immunoblotting with antibodies against two eIF3 subunits (eIF3-b and eIF3-e) failed to detect this complex in our affinity preparations even though these proteins could be readily seen in crude *Arabidopsis* extracts (Fig. 2C and data not shown). Likewise, we failed to detect CSN4 or CSN5 in the affinity preparations, strongly suggesting that the eight-subunit CSN complex evolutionarily related to the RP lid does not form a CP-RP(base)-CSN particle by replacing the lid as proposed (51).

Experiments with wild-type plants indicated that our affinity method also enriched for several proteins nonspecifically (supplemental Table 3). One substantial contaminant at 38 kDa matched the mass of nitrilase, previously shown by us to stick to agarose beads, possibly via its avidity for the aliphatic nitrile groups used to cross-link the resin (40). Its confirmation as nitrilase was subsequently provided by both immunoblot analysis with anti-nitrilase antibodies and by MS/MS sequencing of preparations from both wild-type and *PAG1-FLAG pag1-1* plants (Fig. 2C and supplemental Table 3). HSC70 and Rubisco were also confirmed as contaminants of the affinity procedure by MS and by overexposure of immunoblots prepared with wild-type and *PAG1-FLAG pag1-1* samples and probed with anti-HSC70 and anti-Rubisco large subunit antibodies.

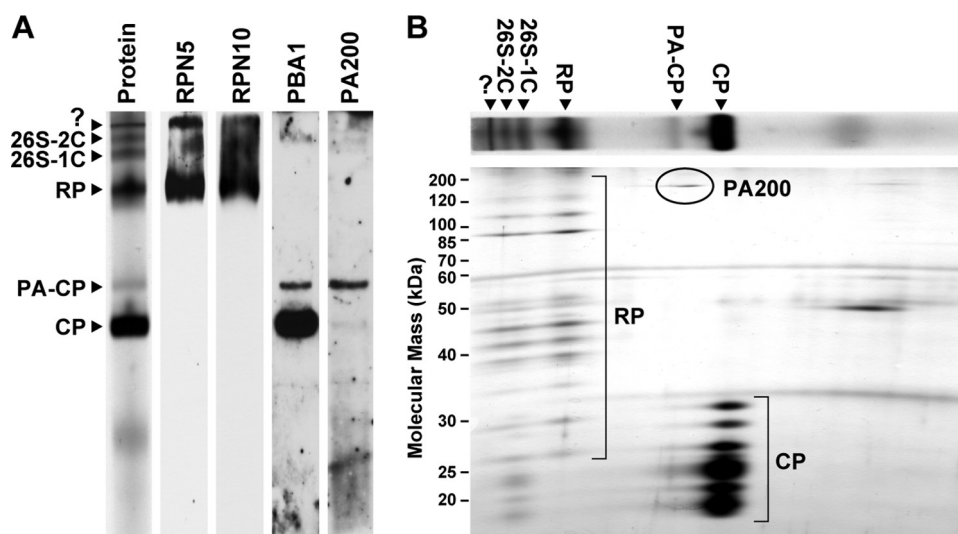
Native PAGE of the affinity preparations followed by immunoblot analyses of representative subunits or by SDS-PAGE in

the second dimension demonstrated that the collection of CP and RP proteins assembled into their respective subcomplexes (Fig. 3, *A* and *B*). In addition, several forms of the 26 S proteasome containing both CP and RP were evident. The fastest mobility form likely represented the CP capped by one RP, whereas the middle form likely represented a CP doubly capped with RP. The slowest migrating form is unknown; intriguingly, it could represent the doubly capped RP-CP complex assembled with a host of other factors as seen for the yeast complex (35). We also detected a new complex of slightly slower mobility as compared with the CP. Subsequent immunoblot analysis and SDS-PAGE in the second dimension revealed that it consists of the CP capped by only PA200 and no RP subunits, and thus it represents a heretofore unknown proteasome type in plants (Fig. 3, *A* and *B*).

**Subunit Composition of *Arabidopsis* 26 S Proteasomes Defined by MS**—We comprehensively determined the subunit composition of the affinity-purified proteasomes using a variety of MS/MS techniques and spectrometers, with a special focus on the presence of the various subunit isoforms and their possible post-translational modifications. The deepest dataset was generated by in-solution trypsinization of unfractionated samples followed by analysis with an LTQ-Orbitrap Velos mass spectrometer using HCD for peptide fragmentation. For a second approach, we separated the polypeptides by SDS-PAGE followed by in-gel trypsinization of gel slices, and we analyzed the peptides with an LCQ-deca XPplus ion-trap mass spectrometer using CID for peptide fragmentation. As a third approach, we digested the polypeptides with LysC in-solution and detected the peptides with the Orbitrap mass spectrometer using ETD for peptide fragmentation. False-positive peptide spectrum matches to the *Arabidopsis* proteome were eliminated upon filtering the data to maintain a false discovery rate below 1%. We also purged proteins identified from wild-type seedlings subjected to the same affinity protocol (e.g. nitrilase, the small and large subunits of Rubisco, and HSC70 (supplemental Table 3)), which presumably bound nonspecifically to the anti-FLAG antibody beads. The combined results for expected RP and CP subunits are summarized in Table 1. Taken together, we confirmed the complete subunit composition of the “core” *Arabidopsis* 26 S proteasome, which includes all 7  $\alpha$  (PAA-PAG) and all 7  $\beta$  (PBA-PBG) subunits of the CP, all 6 RPT subunits, and 11 of the RPN subunits (RPN1–3, RPN5–12, the exception being



## Characterization of the 26 S Proteasome from *Arabidopsis*



**FIGURE 3. Separation of the various proteasome complexes by native PAGE.** *A*, proteasomes affinity-purified from *PAG1-FLAG pag1-1* plants were fractionated by native PAGE in the presence of ATP. Gels were either stained for total protein with silver (*Protein*) or subjected to immunoblot analysis with antibodies against RPN5, RPN10, PBA1, and PA200 (*PA*). Migration of the CP, the PA-CP complex, the RP, and singly and doubly capped 26 S proteasomes (26 S-1C and 26 S-2C) are indicated. The ? identifies a larger complex that could represent the core 26 S proteasome associated with additional factors. *B*, proteasomes separated as in *A* and then subjected to denaturing SDS-PAGE in the second dimension. A representative native PAGE separation is provided above to help orient the complexes. PA200 and the collection of subunits from the RP and CP subcomplexes are located.

RPN13). The MS/MS dataset yielded remarkable sequence coverage for most subunits (42–100%), with a 74% average for all 53 polypeptides (Table 1).

In addition to high sequence coverage, a large number of unique peptides allowed unambiguous identification of nearly every subunit paralog predicted by genome annotations (23, 24, 26, 30, 53). Of the 22 subunits potentially expressed by two *Arabidopsis* genes, we matched specific peptides to both predicted isoforms for all but the following three: PAC2( $\alpha$ 3), RPT1b, and RPN12b (Table 1). The predicted amino acid sequence for PAC2 is highly divergent from that for PAC1. The C-terminal regions of this pair are co-linear but the N-terminal region for PAC2 contains a large insertion of unrelated sequence. The expressed sequence tags for PAC2 only match this insertion, suggesting that the bulk of the locus is not transcribed. Together with the absence of PAC2 protein in the purified 26 S complex, it is highly likely that the PAC2 locus is a pseudogene. The RPN12b locus encodes a substantially truncated protein as compared with RPN12a and lacks expression support (30). Our failure to unambiguously find the RPN12b protein strongly supports the conclusion that this locus is also a pseudogene. We failed to identify the RPT1b isoform but readily detected its paralog RPT1a. Expressed sequence tag data estimate that the RPT1b gene is expressed at a much lower level than RPT1a (5 versus 141 expressed sequence tags (Table 1)). Combined with its low amino acid identity to RPT1a (81%), which is eclipsed only by PAC1/2 and RPN12a/b among 22 proteasome protein pairs in *Arabidopsis*, it is possible that RPT1b is also a pseudogene or, more intriguingly, that it is expressed only under specific conditions.

**Post-translational Modifications of 26 S Proteasome Subunits**—Studies with yeast and mammalian 26 S proteasomes and preliminary analysis of the rice complex showed that a

number of core subunits are or may be affected by various post-translational modifications, including proteolytic processing, acetylation (N-terminal and side chain), myristoylation, and phosphorylation, which may affect their assembly, localization, and/or activity (25, 54–59). In addition, at least one 26 S proteasome subunit (RPN1) may be modified by Ub from the global MS analysis of *Arabidopsis* ubiquitinated proteins (40, 41). Here, we exploited our deep sequence coverage to interrogate more thoroughly the entire *Arabidopsis* proteasome population for such modifications.

Like subunits of yeast and animal 26 S proteasomes (54, 55), a number of *Arabidopsis* subunits are N-acetylated. Of the 53 proteins that compose the core complex, we detected peptides at or close to the predicted initiator methionine for

19. Of these 19, 2 contained an unmodified N-terminal methionine (RPT1a and RPN11); 10 had an acetylated N-terminal methionine, and 7 began with the immediately distal residue (alanine or glycine), which was presumably exposed upon proteolytic removal of the N-terminal methionine (supplemental Table 4). We did not find any nonacetylated versions for the 10 subunits beginning with an N-acetylated methionine, suggesting that this modification is comprehensive. Several subunits were also found to be acetylated on internal residues (PAG1, PAE2, RPN1a, RPT1a, and RPT5a (Table 2)).

Our MS/MS dataset was searched for phosphoryl and myristoyl additions using SEQUEST. Unfortunately, no peptides with supra-threshold values for phosphorylated amino acids were detected. Nor were myristoylated residues identified despite reports that the plant RPT2 subunit bears this modification (25, 60).

Based on studies with the yeast 26 S proteasome, one essential post-translational modification is proteolytic processing at the N terminus of the  $\beta$ 1,  $\beta$ 2, and  $\beta$ 5 subunits (PBA1, PBB1/2, and PBE1/2 in *Arabidopsis*, respectively). This autocatalytic cleavage occurs during CP assembly and exposes an interior threonine whose  $\alpha$ -amino group acts as the nucleophile during peptide bond hydrolysis (61). A search of our MS/MS dataset for PBA1 and PBB1/2 identified peptides that reflect this processing in *Arabidopsis*. In particular, peptides beginning with the expected Thr-13 and Thr-47 residues, respectively (residues 13–31 for PBA1 and 38–58 for PBB1/2), were found in the affinity-purified proteasomes subjected to trypsinization, with no peptides detected for the upstream region (Fig. 4, A–C, and supplemental Fig. 1, A–C).

A similar processing likely occurs for PBE1/2. We failed to detect the tryptic peptides that would reflect this cleavage proximal to Thr-57 (residues 57–66). However, processing

**TABLE 1**  
 MS/MS analysis of core subunits from affinity-purified *Arabidopsis* 26 S proteasomes

20S Core Protease (CP)				Peptide #		# Total Peptides <sup>d</sup>	Sequence Coverage (%) <sup>e</sup>	
Protein	Locus	ESTs <sup>a</sup>	% Identity <sup>b</sup>	Isoform Specific <sup>c</sup>	Ambiguous			
Alpha	PAA1	At5g35590	130	91	15	12	27	78
	PAA2	At2g05840	73		23		35	97
	PAB1	At1g16470	77	99	7	24	31	91
	PAB2	At1g79210	55		2		26	89
	PAC1	At3g22110	110	26	29	2	29	87
	PAC2	At4g15160	65		0		0	0
	PAD1	At3g51260	165	95	11	29	40	96
	PAD2	At5g66140	114		9		38	96
	PAE1	At1g53850	52	98	5	26	31	86
	PAE2	At3g14290	79		2		28	86
	PAF1	At5g42790	126	91	14	27	41	100
	PAF2	At1g47250	34		7		34	93
	PAG1	At2g27020	96			34	34	77
	Beta	PBA1	At4g31300	111			24	24
PBB1		At3g27430	153	97	3	13	16	61
PBB2		At5g40580	37		7		20	55
PBC1		At1g21720	69	96	5	12	17	68
PBC2		At1g77440	62		3		15	71
PBD1		At3g22630	67	91	12	10	22	83
PBD2		At4g14800	112		11		21	98
PBE1		At1g13060	109	97	6	15	21	81
PBE2		At3g26340	32		2		17	62
PBF1		At3g60820	141			33	33	85
PBG1		At1g56450	96			23	23	72
<b>19S Regulatory Particle (RP)</b>								
RPT	RPT1a	At1g53750	141	81	59	8	67	95
	RPT1b	At1g53780	5		0		0	0
	RPT2a	At4g29040	91	99	10	48	58	80
	RPT2b	At2g20140	57		1		49	75
	RPT3	At5g58290	147			50	50	95
	RPT4a	At5g43010	45	96	13	42	55	91
	RPT4b	At1g45000	144		12		54	91
	RPT5a	At3g05530	184	93	30	30	60	94
	RPT5b	At1g09100	43		9		39	67
	RPT6a	At5g19990	144	96	11	52	63	95
RPT6b	At5g20000	88		9		61	94	
RPN	RPN1a	At2g20580	127	90	63	30	93	87
	RPN1b	At4g28470	27		38		68	76
	RPN2a	At2g32730	75	90	19	26	45	54
	RPN2b	At1g04810	34		31		57	66
	RPN3a	At1g20200	83	91	34	24	58	81
	RPN3b	At1g75990	23		17		41	66
	RPN5a	At5g09900	81	92	44	21	58	90
	RPN5b	At5g64760	18		16		41	70
	RPN6	At1g29150	177			43	43	72
	RPN7	At4g24820	169			32	32	69
	RPN8a	At5g05780	155	95	16	13	29	83
	RPN8b	At3g11270	15		5		18	69
	RPN9a	At5g45620	67	93	15	11	26	51
	RPN9b	At4g19006	49		16		27	54
RPN10	At4g38630	109			25	25	42	
RPN11	At5g23540	157			36	36	83	
RPN12a	At1g64520	159	52	29	3	32	96	
RPN12b	At5g42040	0		0		0	0	

<sup>a</sup> Number of expressed sequence tags (ESTs) as of October, 2009.

<sup>b</sup> Percent amino acid identity that was between the full-length protein sequences.

<sup>c</sup> Unique peptides that were identified in the MS/MS dataset.

<sup>d</sup> Total number of unique peptides that were identified in the MS/MS dataset.

<sup>e</sup> Coverage for the PBA1, PBB1/2, and PBE1/2 subunits was calculated based on the sequence of the mature processed protein.



## Characterization of the 26 S Proteasome from *Arabidopsis*

**TABLE 2**

Ubiquitylation of *Arabidopsis* 26 S proteasome subunits

Protein	AGI	Total Pept <sup>a</sup>	Xcorr <sup>b</sup>	Sequence <sup>c</sup>	Modifications <sup>c</sup>
<b>Ubiquitylated Subunits with a Canonical Ub Footprint<sup>d</sup></b>					
PAC1	At3g22110	7	8.12	LYQVEYAMEAIGNAGSAIGLSk*DGVVVLIgek	N-t(Ac), K40(GlyGly)
PAF1	At5g42790	1	5.27	SRSHVVLAcVnK*AQSELSSHQR	C48(Cb), K51(GlyGly), C-t(Me)
PAE1/2	At1g53850/ At3g14290	13	7.08	AIGSGSEGADSSLQEQFNk*DITLQEAETIAVSILK	K185(GlyGly)
		2	6.32	ITSPLEPSSVEK*IMEIDDHIGcAMSGLIADAR	N-t(Ac), K66(GlyGly), C76(Cb)
RPN1a	At2g20580	103	8.91	HNAETEAVDLLMVDLDDLLEHVdk*INFK	K218(GlyGly), T219(Ac)
RPN2a/b	At1g04810/ At2g32730	2	5.14	IDk*LEEAIK	N-t(Ac), K166(GlyGly)
RPN6	At1g29150	11	8.44	IDdk*LLLVDIDLLESKLHFSLR	N-t(Ac), K160(GlyGly)
		7	5.54	IDdk*LLLVDIDLLESK	N-t(Ac), K160(GlyGly)
RPN11	At5g23540	1	6.92	fDTHSk*TNEQTVQEMLSLAAK	N-t(Ac), K238(GlyGly)
RPT1a	At1g53750	4	5.82	KVTek*DFLDAVNK	N-t(Ac), K400(GlyGly)
RPT4a	At5g43010	2	6.04	aIASNIDANFLk*VVSSAIIDKYIGESAR	N-t(Ac), K203(GlyGly)
<b>Ubiquitylated Subunits with a Non-Canonical Ub Footprint<sup>e</sup></b>					
PAG1	At2g27020	63	8.46	LHDEAKDKAFELEMswIcEESk*	S214(Ac), C217(Cb), K221(GlyGly)
		5	6.63	DKAFELEMSwIcEESk*	C217(Cb), S220(Ac), K221(GlyGly)
RPT5a	At3g05530	8	5.75	EKAPcIIIDEIDAIGtk*	C266(Cb), T278(Ac), K279(GlyGly)
		3	7.26	DATEVNHEDFNEGIQVQakk*	K410(GlyGly), C-t(Me)

<sup>a</sup> Total peptides reflect the total number of times the modified peptide was identified in MS/MS data.

<sup>b</sup> Xcorr stringency values were calculated by SEQUEST.

<sup>c</sup> Lowercase letters indicate the amino acids modified by ubiquitylation, acetylation (Ac), methylation (Me), or cysteine carbamidomethylation (Cb). Asterisks locate lysine residues containing the signature Gly-Gly Ub footprint for ubiquitylation. N-t, N terminus, and C-t, C terminus.

<sup>d</sup> Canonical Ub footprint was defined as a lysine with an additional *m/z* of 114 that was not cleaved by trypsin (40).

<sup>e</sup> Non-canonical Ub footprint was defined as a lysine with an additional *m/z* of 114 that was at the end of the peptide following a trypsin cleavage site (42).

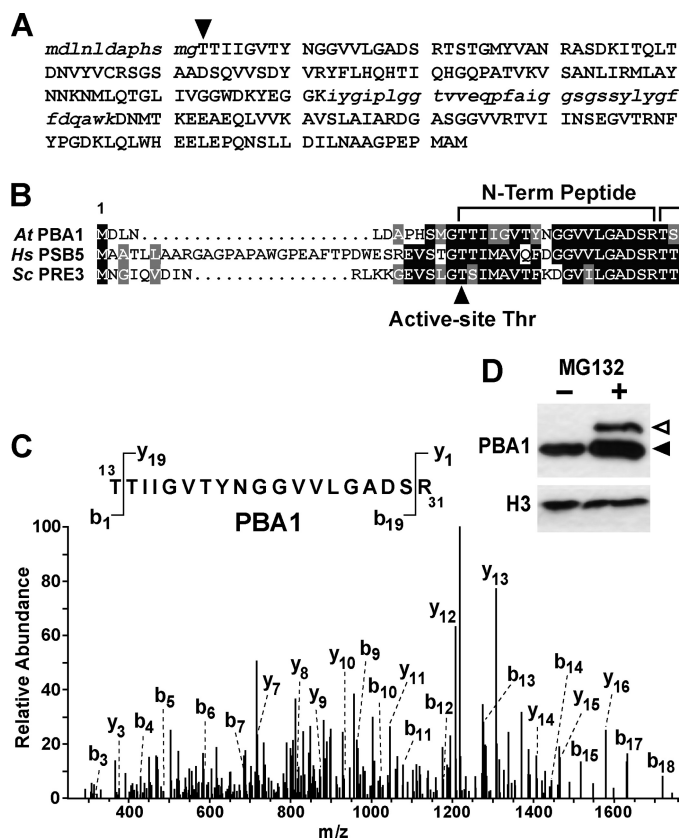
was supported indirectly by the absence of expected peptides upstream of Thr-57 in the PBE1/2 proteins despite extensive sequence coverage of the proteins downstream of this site (64/49% coverage of total sequences or 81%/62% if processed) (supplemental Fig. 1, D and E). Our preliminary catalog of tryptic peptides from PBG1 suggests that it is processed to reveal an N-terminal threonine (Thr-23); this processing is seven residues upstream of the expected position for an active-site threonine and may or may not generate an active peptidase.

Confirmation of  $\beta$ 1 N-terminal processing in *Arabidopsis* was provided by examining the PBA1 protein in plants treated with MG132, which should block this proteolytic maturation (61). Whereas only processed PBA1 was detected immunologically in wild-type plants, a second higher molecular mass form was apparent upon MG132 treatment that matched the expected size of unprocessed PBA1 (23 versus 25 kDa) (Fig. 4D).

Surprisingly, ubiquitylation was the most prevalent modification we found in the affinity-purified proteasomes. Ub addition was detected by its signature tryptic footprint where the target lysine is increased in mass by an isopeptide-linked diglycine derived from the C terminus of Ub (+114 kDa). In most cases, this lysine is immune to trypsin cleavage and thus internal in the peptide (canonical) (40, 41), but in some cases, such cleavage can occur to expose the ubiquitylated lysine at the C terminus (noncanonical) (42). Eleven of the 53 *Arabidopsis* subunits contained a peptide with a high probability Ub footprint (residue of 242.1 *m/z*), nine of the canonical type and two of the noncanonical type (Table 2). The strongest evidence for ubiquitylation was generated for RPN1a; in addition to the unmodified tryptic peptide spanning residues 192–223, we found the

same peptide bearing a canonical Ub footprint at Lys-218, which was abundant in our MS/MS dataset (Fig. 5C). In contrast, Ub footprints were not detected on any RPN1b peptides despite 76% coverage of its sequence, suggesting that the modification is restricted to the  $\alpha$  isoform. Our list of ubiquitylated subunits also included three subunits of the RPT ring, RPN2a/b, RPN6, and four CP  $\alpha$  subunits, including PAG1-FLAG. Some, but not all, of the comparable yeast subunits were also reported to be ubiquitylated from a global analysis of this modification (41).

Further evidence for subunit ubiquitylation was provided by immunoblot analysis of the affinity-purified preparations with anti-Ub antibodies. In addition to trace amounts of Ub trimers and tetramers, and a smear of high molecular mass polyubiquitylated proteins that presumably co-purified with the 26 S proteasome via their association with Ub receptors, we detected two prominent species at 115 and 100 kDa (Fig. 5A). Both their apparent molecular masses and their presence in proteasomes affinity-purified in the presence of ATP but not in the absence strongly suggested that they represent ubiquitylated forms of RPN1a and RPN2a/b (Fig. 5A). The lower molecular mass form was assigned to an RPN1 isoform by probing the SDS-polyacrylamide gels with both anti-Ub and anti-RPN1a antibodies; a 100-kDa species of similar electrophoretic mobility was detected with both antibodies (Fig. 5B). Because the ubiquitylated form of RPN1a should be  $\sim$ 9-kDa larger in mass, this co-migration indicated that either most of the RPN1a pool associated with the *Arabidopsis* 26 S proteasome is modified with Ub or that the appended Ub moiety does not change its electrophoretic mobility.



**FIGURE 4. N-terminal processing of the CP  $\beta$ 1 subunit PBA1.** *A*, amino acid sequence coverage by MS/MS of PBA1 present in proteasomes affinity-purified from *PAG1-FLAG pag1-1* plants. The *uppercase* residues signify those assigned to specific tryptic peptides (85% coverage for the initial translation product). *Lowercase italic* residues signify regions not identified in the peptide sequences. *Arrowhead* locates the most N-terminal residue identified. *B*, amino acid sequence alignment of the N-terminal region of  $\beta$ 1 subunits from *A. thaliana* (At PBA1), *H. sapiens* (Hs PSB5), and *S. cerevisiae* (Sc Pre3). Identical and similar amino acids are shown in the *black* and *gray* boxes, respectively. *Dots* denote gaps. The *arrowhead* marks the N-terminal Thr active site. The *bracket* locates the most N-terminal peptide identified by MS from trypsinized At PBA1. *C*, MS/MS spectrum of the N-terminal peptide of PBA1 (residues 13–31) identified by the *bracket* in *B*. Parent peptide  $MH^+ = 1894.1$  with an *Xcorr* score of 4.59. The locations of signature *b* and *y* ions are indicated along with the calculated peptide sequence. *D*, unprocessed PBA1 accumulates in *Arabidopsis* seedlings treated with MG132. Crude extracts were prepared from wild-type plants with or without a 30-h exposure to 100  $\mu$ M MG132 and then immunoblotted with anti-PBA1 or anti-histone 3 (H3) antibodies (loading control). The *open* and *closed* arrowheads identify the unprocessed and mature forms of PBA1, respectively.

Sequence alignments of RPN1 $\alpha$  and RPN2 $\alpha$ /b revealed that the ubiquitylated lysines are positionally conserved in the *a* but not the *b* isoforms of RPN1 (Lys-218) and in both isoforms of RPN2 $\alpha$ /b (Lys-165) for all plants examined (except the alga *Chlamydomonas reinhardtii*), but substituted for a variety of amino acids in non-plant species (Fig. 5, *E* and *F*). Such limited distributions imply that the ubiquitylation of RPN1 and RPN2 at these sites is restricted to the plant kingdom. When the ubiquitylation sites for the  $\alpha$  subunits were mapped onto the three-dimensional structure of the yeast CP (45), those for PAE1/2 and PAF1 were found to be solvent-exposed and thus accessible to the Ub ligation machinery, whereas those for PAC1 and PAG1 were near the external surface but buried. In the absence of an RP structure, the accessibility of the ubiquitylation sites for the RP subunits is unclear.

**Identification of Proteasome-associated Proteins**—In addition to core RP and CP subunits, we detected a number of proteasome-interacting proteins in our MS/MS dataset (supplemental Table 5). These include Ub, the assembly chaperone PBAC2 (Pba2 in yeast), the accessory factor DSS1/RPN15 (Sem1 in yeast), the UPB16 de-ubiquitylating enzyme (Ubp8/Usp42 in yeast/humans), and the CP activator PA200, which was also detected immunologically (Figs. 1*D* and 2*C*). A collection of proteins unique to plants was also present in the dataset. Confirmation that they actually associate specifically with *Arabidopsis* proteasomes is underway.

In a previous study, we detected immunologically in our proteasome preparations members of the RAD23 and DSK2 family, which transiently interact with the 26 S proteasome and serve as shuttles to deliver ubiquitylated substrates (18). However, none of these shuttle proteins was detected here by MS, suggesting that they interact at substoichiometric levels with the complex. We also failed to positively identify several proteasome-associated proteins first identified in yeast, including the Ub receptor RPN13, UBP6, UPL7 (yeast Hul5), and ECM29 (2, 17), even though genes encoding obvious orthologs are present in *Arabidopsis*. The absence of these proteins in the MS/MS dataset implies that either their interactions are too weak to even survive the mild isolation conditions used here, the bound forms of the proteins are present at very low levels, or that they do not interact with the *Arabidopsis* particle. Neither searches of our MS/MS dataset nor BLAST analysis of the *Arabidopsis* genome found obvious orthologs of the  $\alpha$  or  $\beta$  subunits comprising the mammalian heterohexameric PA26/11S regulator (19), suggesting that plants do not use this alternative RP in proteasomal control.

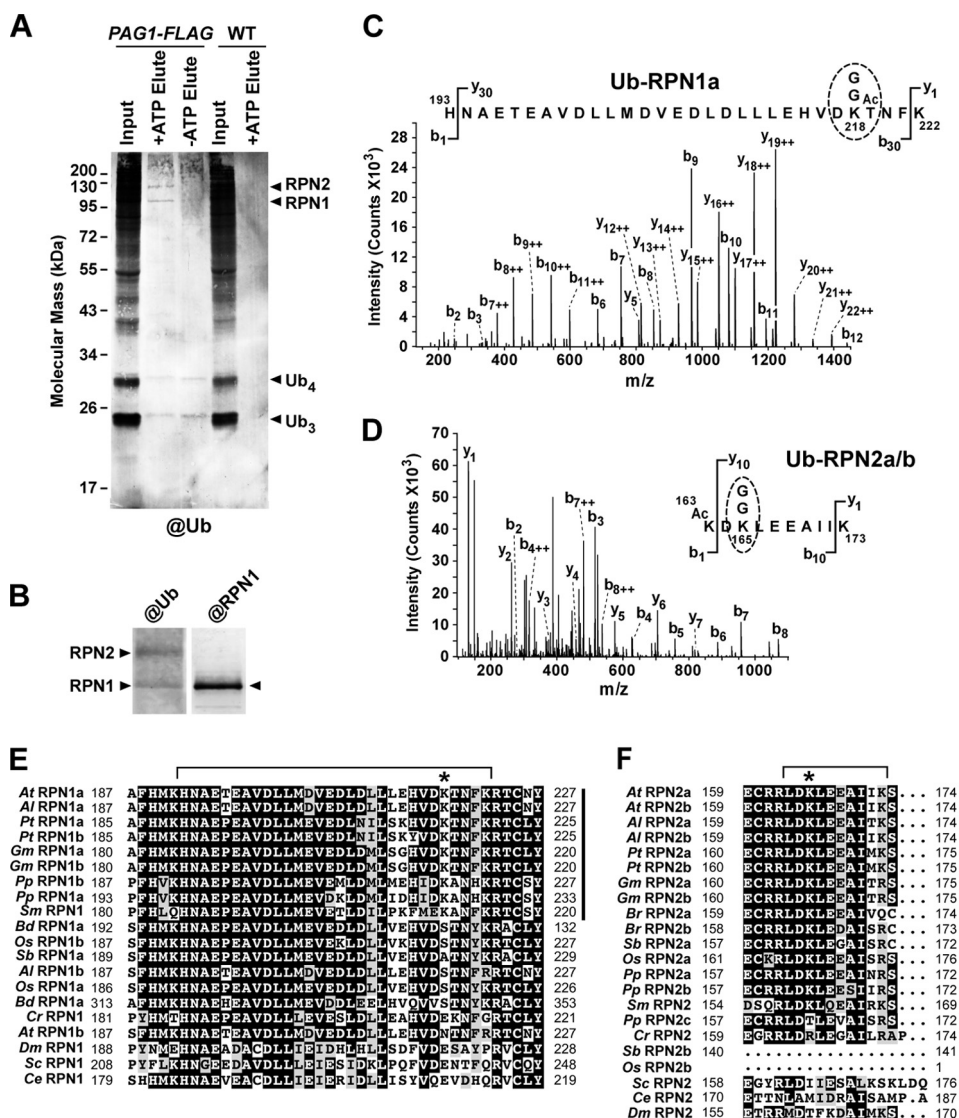
**Genetic Analysis of PA200 in *Arabidopsis***—The most intriguing proteasome-interacting protein was PA200, given our detection of the PA200-CP complex by native PAGE in the affinity-purified preparations or even following glycerol gradient centrifugation of crude *Arabidopsis* extracts (Figs. 3 and 1*C*) and the fact that it could represent an alternative proteasome in plants. The mammalian and yeast versions stimulate the peptidase activity of the CP, likely by opening the gate, and may participate in the Ub-independent degradation (17, 20, 22). *Arabidopsis* PA200 shares only 22% amino acid sequence identity with both its yeast and human counterparts but can be easily recognized by the organization of HEAT repeats, which help form its signature solenoid structure (Fig. 6*A*) (20–22).

To better understand the functions of PA200 in plants, we developed a collection of T-DNA insertion mutants that potentially block expression of the single *Arabidopsis* PA200 gene (Fig. 6*A*). Both RT-PCR and immunoblot analysis with anti-PA200 antibodies documented that two of the mutants *pa200-2* and *pa200-3* failed to accumulate both the full-length PA200 mRNA and protein and thus likely represent null alleles (Fig. 6, *B* and *C*).

Surprisingly, the *pa200* mutants were phenotypically indistinguishable to wild type under a variety of growth conditions, indicating that this activator is not essential in *Arabidopsis* (Fig. 6, *D* and *E*). These conditions included darkness, short and long day photoperiods, continuous red or far-red light, exposure to



## Characterization of the 26 S Proteasome from Arabidopsis



**FIGURE 5. RPN1a and RPN2 assembled into the 26 S proteasome are ubiquitylated.** **A**, extracts from 10-day-old WT and *PAG1-FLAG pag1-1* plants were subjected to the affinity purification method for proteasomes in the presence or absence of ATP (see Fig. 2A). The input and eluted fractions were subjected to SDS-PAGE and immunoblot analysis with anti-Ub antibodies. Arrowheads identify poly-Ub trimers and tetramers and the approximate migration position of RPN1 and RPN2. **B**, co-migration of RPN1 with a ubiquitylated species in affinity-purified proteasomes. Adjacent gel lanes were subjected to immunoblot analysis with anti-Ub or anti-RPN1 antibodies. The arrowheads identify RPN1 and the approximate migration positions of RPN2. **C**, ESI-MS/MS spectrum of a tryptic peptide from RPN1a ( $MH^+ = 3636.78$  m/z) that contains a Gly-Gly footprint derived from Ub attached to Lys-218. **D**, ESI-MS/MS spectrum of a tryptic peptide from RPN2a/b ( $MH^+ = 1327.80$  m/z) that contains a Gly-Gly footprint derived from Ub attached to Lys-165. **E** and **F**, amino acid sequence comparison of RPN1 (**E**) and RPN2 (**F**) from various eukaryotes at the region surrounding the ubiquitylation sites identified in the *Arabidopsis* proteins. Identical and similar residues are shown in black and gray boxes, respectively. Dots denote gaps. *A. thaliana*, At; *A. lyrata*, Al; *G. max*, Gm; *P. trichocarpa*, Pt; *O. sativa*, Os; *Sorghum bicolor*, Sb; *Brachypodium distachyon* Bd; *S. moellendorffii*, Sm; *C. reinhardtii*, Cr; *P. patens*, Pp; *Caenorhabditis elegans*, Ce; *D. melanogaster*, Dm; and *S. cerevisiae*, Sc. The bracket locates the ubiquitylated peptide detected by ESI-MS/MS with the asterisk marking the Lys containing the Ub footprint (Gly-Gly). The vertical line identifies sequences containing the conserved ubiquitylated Lys.

MG132, which should block CP activity, and growth in the presence of amino acid analogs, which should increase the abundance of abnormal proteins that require proteasomal removal (53). The levels of Ub, free poly-Ub chains, and Ub-protein conjugates were also unaffected in the *pa200* backgrounds ( $\pm$ MG132) as compared with wild type, suggesting that the PA200-CP complex works independently of ubiquitylation (Fig. 6F). Whereas yeast *blm10* mutants are modestly hypersensitive to the DNA-damaging agent bleomycin (21),

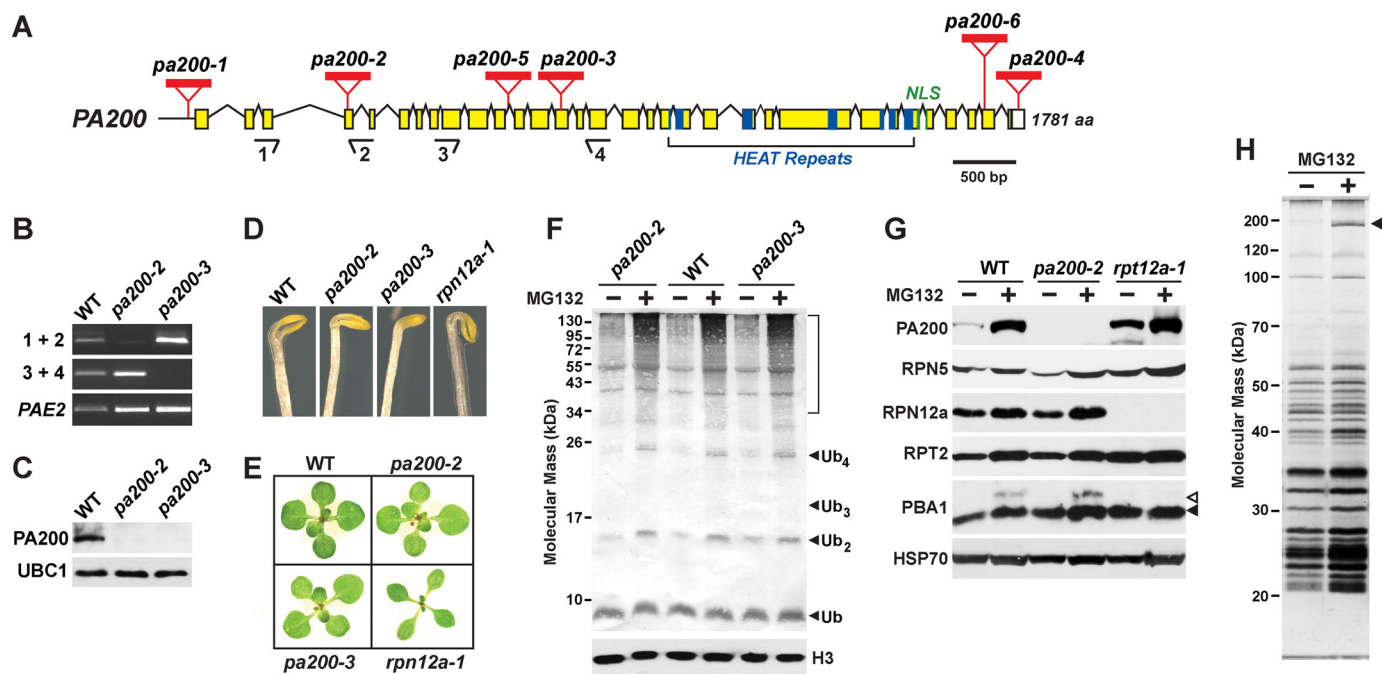
*Arabidopsis pa200-2* and *pa200-3* seedlings were not more sensitive to this drug.

The only response we observed for PA200 was a substantial increase in the protein upon exposure of *Arabidopsis* seedlings to MG132 or in mutant backgrounds that dampen RP activity (e.g. *rpn12a-1* (30)) (Fig. 6G). This rise implies that PA200 is either a proteasome substrate or more likely that PA200, like subunits of the CP and RP (26), is part of a regulatory system that coordinately responds to proteolytic demand. The increase in PA200 resulted in the assembly of more PA200-CP complexes. As compared with proteasome preparations affinity-purified via the *PAG1-FLAG* protein from non-treated plants, those affinity purified after MG132 treatment contained substantially more PA200 (Fig. 6H).

**Phylogenetic Analysis of Proteasome Subunits in Plants**—The proteasomal incorporation of both isoforms for most CP and RP subunits encoded by gene pairs indicates that the *Arabidopsis* complex is actually a heterogeneous collection of particles assembled with an assortment of isoform combinations. This heterogeneity in turn raises the possibility some pairs have subfunctionalized to generate distinct 26 S proteasome types. Recent analyses of RPN1a/b, RPN5a/b, and RPT2a/b support such nonredundancy by showing that the subunits pairs have unique expression patterns and possibly both common and nonoverlapping activities (13, 27, 29, 31, 32).

As a further test, we examined phylogenetically how several CP and RP paralogs evolved with the prediction that the earlier these duplicates appeared during plant evolution and the more they were retained in the different plant lineages, the more likely it would be that the paralogs acquired distinct functions. Our analyses included CP (PAA1/2) and RP (RPN1a/b, RPN2a/b, RPT1a/b) protein pairs from a number of eudicots (including *A. thaliana* and *Arabidopsis lyrata*), monocots, the lower plants *Selaginella moellendorffii* and *P. patens*, and the alga *C. reinhardtii*. The phylogenetic trees were determined via Bayesian estimation and were rooted with obvious non-plant orthologs (Fig. 7 and supplemental Fig. 2).





**FIGURE 6. Molecular and genetic analyses of Arabidopsis PA200.** *A*, diagram of the Arabidopsis PA200 gene. Yellow and white boxes denote coding and 3'-untranslated region exons, respectively. Lines show introns. Coding regions for the HEAT repeats and nuclear localization signal (NLS) are in blue and green, respectively. T-DNA insertion sites for the *pa200-1-6* mutants are shown in red. Arrows locate the positions of the primers used for RT-PCR analysis in *B*. *aa*, amino acids. *B*, RT-PCR analysis of 2-week-old *pa200-2* and *pa200-3* seedlings. Total RNA isolated from WT and mutant seedlings was subjected to RT-PCR using the PA200 primers located by the arrows in *A*. A primer pair specific to PAE2 was used as an internal control. *C*, immunoblot analyses of total protein from 1-week-old WT and *pa200-2* and *pa200-3* mutant seedlings with anti-PA200 and anti-UBC1 (loading control) antibodies. *D*, 10-day-old etiolated *pa200-2* and *pa200-3* seedling as compared with WT and *rpn12a-1* seedlings. *E*, 12-day-old green WT, *pa200-2*, *pa200-3*, and *rpn12a-1* seedlings grown under a long-day photoperiod (16 h light/8 h dark). *F*, immunoblot analysis of Ub-conjugate levels. Seedlings were either treated with DMSO or 100  $\mu$ M MG132 dissolved in DMSO. Equivalent amounts of total protein were subjected to SDS-PAGE and immunoblotted with anti-Ub antibodies and confirmed by probing with anti-histone 3 (H3) antibodies. Arrowheads indicated free Ub, Ub dimers, trimers, and tetramers. The bracket indicates Ub conjugates. *G*, increased PA200 protein levels in response to proteasome inhibition. Four-day-old WT, *pa200-2*, and *rpn12a-1* seedlings were treated for 30 h with DMSO or 100  $\mu$ M MG132 dissolved in DMSO. Equivalent amounts of total protein were subjected to SDS-PAGE and immunoblotted with antibodies against PA200, RPN5a, RPN12a, RPT2a, and PBA1, and HSP70 (loading control). The open and closed arrowheads identify the unprocessed and mature forms of PBA1, respectively. *H*, PAG1-FLAG affinity purification of proteasomes from 10-day-old seedlings treated with or without 100  $\mu$ M MG132 for 30 h. The arrowhead locates PA200, the identity of which was confirmed by MS/MS analysis of the excised gel slice.

Comparisons of the trees revealed that the duplication of the plant CP and RP genes likely occurred not in a common ancestor but at different times during plant evolution, depending on the lineage branch. Overall, the protein pairs grouped according to the main plant lineages (eudicot, monocot, lower plant, and alga). By and large, the eudicot proteins clustered with their species paralog and separate from the orthologous pair in other eudicots, implying that each pair arose separately by relatively recent species-specific duplication events. Likely causes were the whole scale genome duplications seen for many plant species such as *A. thaliana* (62). In contrast, individuals from the monocot pairs commonly clustered with orthologs from other monocots and not with their paralogs, implying that these pairs appeared early during monocot evolution and were retained as the modern genomes developed. The only deviation from this pattern was RPT1, where the RPT1b isoform in both *A. thaliana* and *A. lyrata* formed a separate subclade away from the other eudicot and monocot sequences (Fig. 7B). Assuming that the RPT1b locus is a pseudogene in both *A. thaliana* and *A. lyrata*, this distinct subclade likely reflects the diminished pressure to conserve coding sequence after gene inactivation.

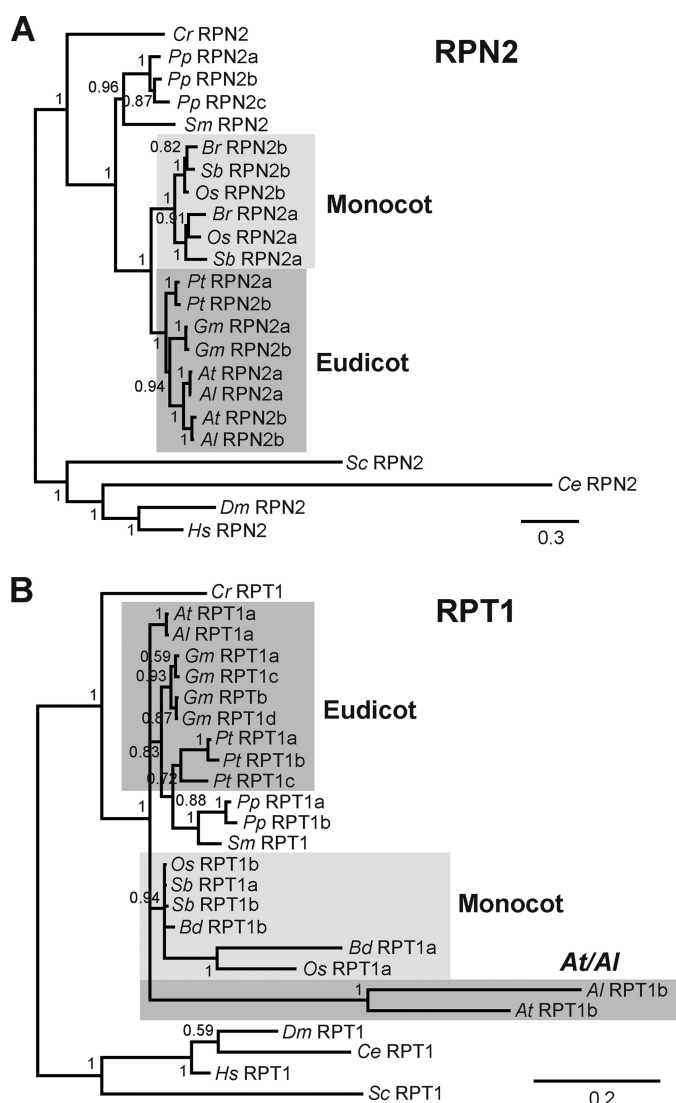
Taken together, it appears unlikely that plants share a common set of proteasome subunit isoforms, strongly suggesting that if isoform-specific functions do exist, they are not shared

among all plants. Furthermore, the more recent evolutionary origins of the eudicot paralogs, as compared with those in monocots, imply that any nonredundant functions that do exist in eudicots should be species-specific.

## DISCUSSION

Given its central position within the Ub/26 S proteasome system, the 26 S proteasome assumes a major role in the selective removal of both aberrant polypeptides and important short lived regulators. Based on estimates that plants such as *Arabidopsis* express >1000 different Ub-protein ligases that ubiquitylate an equal or greater number of targets (1, 3), it is expected that this self-compartmentalized protease is uniquely designed to operate with high selectivity for the appended Ub moiety but with low specificity for the sequence of the modified target. To provide insights into the organization and functions of the plant 26 S proteasome, we present here a comprehensive description of the particle from young *Arabidopsis* seedlings, using a robust affinity method for its isolation. MS and biochemical analyses of the resulting preparations revealed that plants actually assemble a heterogeneous collection of proteasomes that may play important roles in handling the myriad of potential targets that are committed by both Ub-dependent and -independent mechanisms.

## Characterization of the 26 S Proteasome from *Arabidopsis*



**FIGURE 7. Phylogenetic analysis of the proteasome subunits RPN2 (A) and RPT1 (B).** Bayesian phylogenetic analyses were performed using protein sequences from *A. thaliana* (At), *A. lyrata* (Al), *G. max* (Gm), *P. trichocarpa* (Pt), *O. sativa* (Os), *S. bicolor* (Sb), *B. distachyon* (Bd), *S. moellendorffii* (Sm), *C. reinhardtii* (Cr), *P. patens* (Pp), *C. elegans* (Ce), *D. melanogaster* (Dm), *H. sapiens* (Hs), and *S. cerevisiae* (Sc). Eudicot- and monocot-specific subclades are highlighted by the shaded boxes. The unique At/Al subclade for RPT1b is also highlighted. Posterior probabilities were calculated from 1000 generations, and values are indicated at each branch node. Trees were rooted with obvious non-plant orthologs. The scale bar represents the average number of substitutions per site.

Central to our affinity method was the genetic replacement of a CP subunit with an affinity-tagged variant followed by the use of a single, mild chromatography step based on the tag to rapidly isolate the complex directly from crude extracts. For simplicity, we focused on the PAG1 subunit given that it is encoded by a single *Arabidopsis* gene. In theory, many other subunits should also be amenable if they can be engineered with solvent-exposed tags.

Like various subunits of the *Arabidopsis* RP lid and base (13, 27–29, 31), our genetic analysis of PAG1 revealed that the CP is also essential in *Arabidopsis*. Interestingly, a common theme of these CP and RP mutants is that they block male but not female gametogenesis. This block could reflect a special role for the

26 S proteasome during male meiosis as seen in an *Arabidopsis* deletion of the pollen-specific E3 component UBL17 (63), or more likely, it could reflect the unique development program that delivers the sperm cells to the ovule during fertilization. Although haploid egg cells missing proteasome subunits may still have sufficient levels of the 26 S proteasome provided by the maternal diploid megasporocyte, the extensive growth of the haploid pollen tube cell as it delivers the two sperm cells may eventually dilute proteasome levels below a viable threshold.

Here, we affinity-captured proteolytic complexes sharing the CP subunit PAG1, using young *Arabidopsis* seedlings as the source. MS/MS analyses revealed that all but three of the potential subunit paralogs based on the *Arabidopsis* genome are incorporated into the particle assembled in this tissue, except for the predicted products from the *PAC2*, *RPN12b*, and *RPT1b* loci. Consequently, we provide further support for the proposal that plants actually contain a large array of compositionally distinct proteasomes (26). They could be assembled at random to generate a wide assortment of subtly different particles or could be deliberately constructed by integrating together specific isoforms to create a more defined subset. Such deliberate matching could imply that functionally unique plant proteasomes exist. Such a scenario is exemplified by the mammalian immunoproteasome, which replaces the  $\beta$ 1,  $\beta$ 2, and  $\beta$ 5 catalytic subunits of the CP with specialized isoforms that release peptides more suitable for antigen presentation (64). Work is now underway to test this possibility by affinity-purifying proteasomes containing tags appended to specific isoforms of individual CP and RP subunits.

Phylogenetic analyses of the 26 S proteasome within the plant kingdom strongly suggest that the various paralogs arose not from a common ancestor but were generated by multiple independent duplications in the various plant lineages. In particular, we predict that the monocot isoforms arose by genome duplication(s) in a common ancestor after the eudicot/monocot split, whereas the eudicot isoforms separately arose later by species-specific duplication events. Consequently, if subfunctionalization of paralogs has occurred, its consequences may not pervade the entire plant kingdom but be more restricted to individual species. The most likely scenario is subfunctionalization of expression patterns, which has already been observed between several *Arabidopsis* RP paralogs (13, 28, 31, 65).

For *PAC2*, *RPT1b*, and *RPN12b* loci for which protein products were not detected, the derived amino acid sequences have the lowest identity as compared with their incorporated paralog within the collection of *Arabidopsis* RP and CP proteasome subunits. The predicted *PAC2* and *RPN12b* proteins are also substantially truncated, making it highly likely that both are pseudogenes. For *RPT1b*, its activity is unclear. The corresponding protein retained a high sequence identity to its RPT1a paralog (81%), but both it and its *A. lyrata* ortholog are clearly distinct phylogenetically from others in the plant RPT1 family. Its significantly lower expression in *Arabidopsis* combined with its absence in our MS/MS dataset suggest that RPT1b is either not expressed, has tissue/developmental stage-restricted expression patterns, or more remotely, is not assembled into a complex that binds the CP.

In-depth MS analyses revealed that the *Arabidopsis* 26 S proteasome, despite its heterogeneity, is remarkably analogous to the yeast and human versions (35, 55). In particular, we found that the PBA1( $\beta$ 1) and PBB1/2( $\beta$ 2) subunits and likely also the PBE1/2 ( $\beta$ 5) subunit are post-translationally processed to reveal an N-terminal threonine that we presume becomes part of their active sites. Although not conclusive, the sensitivity of PBA1 cleavage to MG132 implies that the CP itself is the responsible protease as demonstrated genetically in yeast (61).

Like the yeast and mammalian subunits, the N terminus of a number of other *Arabidopsis* subunits is altered by N-acetylation of the initiator methionine or by N-terminal processing (supplemental Table 4). The roles of these modification/processing events in proteasome assembly/function are not yet clear. Although several studies with non-plant proteasomes identified various subunits that are modified by phosphorylation, we failed to detect this modification using ETD MS/MS methods in any of the *Arabidopsis* subunits. Final conclusions about phosphorylation status will ultimately require more sensitive methods to detect phosphoproteins, including the use of immobilized metal chelate chromatography to enrich for phosphopeptides.

The most surprisingly finding was the extensive modification of the *Arabidopsis* 26 S proteasome subunits by Ub. Prior MS/MS studies of global ubiquitylation detected a number of yeast subunits containing one or more Ub moieties, including Pre3( $\beta$ 1), Pup1( $\beta$ 2), Pre2( $\beta$ 5), RPN1, RPN2, RPN3, RPN5, RPN6, RPN7, RPN8, RPN9, and RPN13 (41), whereas similar studies by us in *Arabidopsis* detected ubiquitylated RPN1 (40). Neither study identified the site(s) of Ub addition nor was it clear whether the subunits were ubiquitylated while bound to the 26 S complex. Here, we identified Ub footprints on 11 proteasome subunits, and based on our method of enrichment, most of these subunits should contain the Ub appendage while incorporated into the 26 S proteasome.

The only exception might be the ubiquitylated form of PAG1 (as PAG1-FLAG), which could have been purified in the free form in addition to that assembled into proteasomes. In fact, the frequency with which we detected ubiquitylated PAG1 in *PAG1-FLAG pag1-1* plants as compared with other subunits (Table 2) raises the possibility that this modification reflects the turnover of excess PAG1-FLAG, which failed to assemble into the CP. Such turnover has been reported previously for *Arabidopsis* RPN10 (53), suggesting that unassembled proteasome subunits are rapidly removed *in planta*.

Immunoblot analysis of the 26 S proteasome preparations with anti-Ub antibodies suggested that RPN1 and RPN2 are the most abundant ubiquitylated proteins in the *Arabidopsis* complex (at least in young seedlings). Although sequence alignments surrounding the Ub attachment sites (Lys-218 in RPN1a and Lys-165 in RPN12/a/b) imply that this modification is restricted to plants, the detection of ubiquitylated yeast RPN1 could indicate that the modification is more widespread (41).

What are the functions/fates of ubiquitylated 26 S proteasomes? One possibility is that they reflect the process whereby plants and other eukaryotes remove unwanted or nonfunctional complexes. Through ubiquitylation and subsequent subunit extraction, this 2.5-MDa complex could be processively

disassembled followed by Ub-mediated degradation of the individual free subunits. Alternatively, Ub addition could impart new functions to specific subunits. For example, given the positions of RPN1 and RPN2 in the RP base close to its interface with the lid (49), and the proposed role of RPN1 as a Ub receptor (66), the ubiquitylation of these proteins could have important implications in lid/base binding and/or the association of ubiquitylated substrates with the RP. Contrasting possibilities are that ubiquitylation of RPN1 either blocks subsequent docking of Ub-protein conjugates by interacting with its own Ub recognition domain or that it promotes docking of conjugates via the interaction of the Ub moieties with the Ub-binding motifs of shuttle proteins such as RAD23, which deliver ubiquitylated cargo to the particle (18, 66). Clearly, the analysis of RPN1 and RPN2 mutants that inhibit Ub addition (*e.g.* arginine substitutions at Lys-218 and Lys-165, respectively) are needed to clarify its importance.

In addition to core CP and RP subunits, we identified a number of accessory proteins, including the PBAC2 (Pba2 in yeast) chaperonin, whose ortholog in yeast helps with assembly of the CP/RP holo-complex, DSS1/RPN15 (Sem1 in yeast), which may alter proteasome subunit composition and/or localization, and the deubiquitylating enzyme UBP16 (2, 17, 67). The closest yeast and human orthologs of UBP16 are Ubp8 and Usp42, which have been implicated in histone H2B deubiquitylation by the SAGA complex and embryogenesis, respectively (68, 69). Whether UBP16 represents a functional analog of these DUBs remains to be determined. As opposed to the yeast particle, we failed to detect the Ub receptor RPN13, ECM29, UBP6, and UPL7(Hul5) despite the presence of obvious orthologs in the *Arabidopsis* genome. The absence of these potential cofactors could indicate that the affinity-purified proteases characterized here are still missing important accessory proteins despite the mild conditions used for isolation. Retaining association of these cofactors could require chemical cross-linking before tissue homogenization to stabilize their weak affinity.

Another accessory protein associated with the *Arabidopsis* proteasomes is PA200 that forms a complex with the CP. This alternative CP activator has been implicated in Ub-independent protein breakdown, but its exact physiological role(s) remain unclear (20, 21). Like that in yeast, PA200 is not essential in *Arabidopsis*. In fact, null *pa200* mutants were indistinguishable to wild type under a variety of growth conditions. That the level of PA200 in *Arabidopsis* seedlings substantially increases in response to conditions that limit 26 S proteasome activity (MG132 treatment or the *rpn12a-1* mutant background) could indicate that PA200 is a proteasome substrate. Alternatively, observations that MG132 also markedly increases the levels of both the *PA200* mRNA<sup>5</sup> and the PA200-capped CP complex (Fig. 6H) raise the intriguing possibility that the PA200-CP proteasome represents a stress-induced proteolytic particle. Hopefully, the analysis of *pa200* mutants in combination with 26 S proteasome RP mutants may help delineate the function(s) of the PA200-CP complex, especially under suboptimal growth conditions or situations that stress 26 S proteasome capacity.

<sup>5</sup> K. H. Lee and R. D. Vierstra, unpublished data.



## Characterization of the 26 S Proteasome from Arabidopsis

In summary, analysis of our affinity-purified preparations has revealed that *Arabidopsis* and likely other plants, assemble a heterogeneous collection of proteasomes. By exploiting this purification strategy, it should now be possible to track the abundance of the various complexes and subcomplexes and their post-translational modifications in specific tissues and developmental states and, under various growth conditions, to reveal how each contributes to the architecture and functions of these proteolytic machines.

*Acknowledgments*—We thank Joseph Walker, Dr. Maria Hansson, and David Gemperline for technical help; Drs. Archie Portis, Xing-Wang Deng, Daniel Chamovitz, Mark Estelle, and Charles Guy for various antibodies; and the Arabidopsis Biological Resource Center for providing the expressed sequence tag clones and SALK T-DNA lines.

### REFERENCES

- Vierstra, R. D. (2009) *Nat. Rev. Mol. Cell Biol.* **10**, 385–397
- Finley, D. (2009) *Annu. Rev. Biochem.* **78**, 477–513
- Smalle, J., and Vierstra, R. D. (2004) *Annu. Rev. Plant Biol.* **55**, 555–590
- Fu, H., Reis, N., Lee, Y., Glickman, M. H., and Vierstra, R. D. (2001) *EMBO J.* **20**, 7096–7107
- Rabl, J., Smith, D. M., Yu, Y., Chang, S. C., Goldberg, A. L., and Cheng, Y. (2008) *Mol. Cell* **30**, 360–368
- Groll, M., Bajorek, M., Köhler, A., Moroder, L., Rubin, D. M., Huber, R., Glickman, M. H., and Finley, D. (2000) *Nat. Struct. Biol.* **7**, 1062–1067
- Elsasser, S., Chandler-Militello, D., Müller, B., Hanna, J., and Finley, D. (2004) *J. Biol. Chem.* **279**, 26817–26822
- Husnjak, K., Elsasser, S., Zhang, N., Chen, X., Randles, L., Shi, Y., Hofmann, K., Walters, K. J., Finley, D., and Dikic, I. (2008) *Nature* **453**, 481–488
- Schreiner, P., Chen, X., Husnjak, K., Randles, L., Zhang, N., Elsasser, S., Finley, D., Dikic, I., Walters, K. J., and Groll, M. (2008) *Nature* **453**, 548–552
- van Nocker, S., Sadis, S., Rubin, D. M., Glickman, M., Fu, H., Coux, O., Wefes, I., Finley, D., and Vierstra, R. D. (1996) *Mol. Cell Biol.* **16**, 6020–6028
- Saeki, Y., Sone, T., Toh-e, A., and Yokosawa, H. (2002) *Biochem. Biophys. Res. Commun.* **296**, 813–819
- Verma, R., Aravind, L., Oania, R., McDonald, W. H., Yates, J. R., 3rd, Koonin, E. V., and Deshaies, R. J. (2002) *Science* **298**, 611–615
- Book, A. J., Smalle, J., Lee, K. H., Yang, P., Walker, J. M., Casper, S., Holmes, J. H., Russo, L. A., Buzzinotti, Z. W., Jenik, P. D., and Vierstra, R. D. (2009) *Plant Cell* **21**, 460–478
- Hirano, Y., Hendil, K. B., Yashiroda, H., Iemura, S., Nagane, R., Hioki, Y., Natsume, T., Tanaka, K., and Murata, S. (2005) *Nature* **437**, 1381–1385
- Funakoshi, M., Tomko, R. J., Jr., Kobayashi, H., and Hochstrasser, M. (2009) *Cell* **137**, 887–899
- Roelofs, J., Park, S., Haas, W., Tian, G., McAllister, F. E., Huo, Y., Lee, B. H., Zhang, F., Shi, Y., Gygi, S. P., and Finley, D. (2009) *Nature* **459**, 861–865
- Schmidt, M., Hanna, J., Elsasser, S., and Finley, D. (2005) *Biol. Chem.* **386**, 725–737
- Farmer, L. M., Book, A. J., Lee, K. H., Lin, Y. L., Fu, H., and Vierstra, R. D. (2010) *Plant Cell* **22**, 124–142
- Förster, A., Masters, E. L., Whitby, F. G., Robinson, H., and Hill, C. P. (2005) *Mol. Cell* **18**, 589–599
- Iwanczyk, J., Sadre-Bazzaz, K., Ferrell, K., Kondrashkina, E., Formosa, T., Hill, C. P., and Ortega, J. (2006) *J. Mol. Biol.* **363**, 648–659
- Schmidt, M., Haas, W., Crosas, B., Santamaria, P. G., Gygi, S. P., Walz, T., and Finley, D. (2005) *Nat. Struct. Mol. Biol.* **12**, 294–303
- Ustrell, V., Hoffman, L., Pratt, G., and Rechsteiner, M. (2002) *EMBO J.* **21**, 3516–3525
- Fu, H., Doelling, J. H., Arendt, C. S., Hochstrasser, M., and Vierstra, R. D. (1998) *Genetics* **149**, 677–692
- Fu, H., Doelling, J. H., Rubin, D. M., and Vierstra, R. D. (1999) *Plant J.* **18**, 529–539
- Shibahara, T., Kawasaki, H., and Hirano, H. (2002) *Eur. J. Biochem.* **269**, 1474–1483
- Yang, P., Fu, H., Walker, J., Papa, C. M., Smalle, J., Ju, Y. M., and Vierstra, R. D. (2004) *J. Biol. Chem.* **279**, 6401–6413
- Brukhin, V., Gheyselsinck, J., Gagliardini, V., Genschik, P., and Grossniklaus, U. (2005) *Plant Cell* **17**, 2723–2737
- Gallois, J. L., Guyon-Debast, A., Lécureuil, A., Vezon, D., Carpentier, V., Bonhomme, S., and Guerche, P. (2009) *Plant Cell* **21**, 442–459
- Kurepa, J., Wang, S., Li, Y., Zaitlin, D., Pierce, A. J., and Smalle, J. A. (2009) *Plant Physiol.* **150**, 178–189
- Smalle, J., Kurepa, J., Yang, P., Babiychuk, E., Kushnir, S., Durski, A., and Vierstra, R. D. (2002) *Plant Cell* **14**, 17–32
- Sonoda, Y., Sako, K., Maki, Y., Yamazaki, N., Yamamoto, H., Ikeda, A., and Yamaguchi, J. (2009) *Plant J.* **60**, 68–78
- Ueda, M., Matsui, K., Ishiguro, S., Sano, R., Wada, T., Paponov, I., Palme, K., and Okada, K. (2004) *Development* **131**, 2101–2111
- Fujinami, K., Tanahashi, N., Tanaka, K., Ichihara, A., Cejka, Z., Baumeister, W., Miyawaki, M., Sato, T., and Nakagawa, H. (1994) *J. Biol. Chem.* **269**, 25905–25910
- Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, J., and Deshaies, R. J. (2000) *Mol. Biol. Cell* **11**, 3425–3439
- Leggett, D. S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R. T., Walz, T., Ploegh, H., and Finley, D. (2002) *Mol. Cell* **10**, 495–507
- Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C. S. (2006) *Plant J.* **45**, 616–629
- Perry, R. H., Cooks, R. G., and Noll, R. J. (2008) *Mass Spectrom. Rev.* **27**, 661–699
- Wiesner, J., Premsler, T., and Sickmann, A. (2008) *Proteomics* **8**, 4466–4483
- Wenner, B. R., and Lynn, B. C. (2004) *J. Am. Soc. Mass Spectrom.* **15**, 150–157
- Saracco, S. A., Hansson, M., Scalf, M., Walker, J. M., Smith, L. M., and Vierstra, R. D. (2009) *Plant J.* **59**, 344–358
- Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S. P. (2003) *Nat. Biotechnol.* **21**, 921–926
- Denis, N. J., Vasilescu, J., Lambert, J. P., Smith, J. C., and Figeys, D. (2007) *Proteomics* **7**, 868–874
- Rohrbough, J. G., Breci, L., Merchant, N., Miller, S., and Haynes, P. A. *J. Biomol. Technol.* **17**, 327–332
- Huelsenbeck, J. P., and Dyer, K. A. (2004) *J. Mol. Evol.* **58**, 661–672
- Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D., and Huber, R. (1997) *Nature* **386**, 463–471
- Fujiwara, T., Tanaka, K., Orino, E., Yoshimura, T., Kumatori, A., Tamura, T., Chung, C. H., Nakai, T., Yamaguchi, K., Shin, S., et al. (1990) *J. Biol. Chem.* **265**, 16604–16613
- Kleijnjen, M. F., Roelofs, J., Park, S., Hathaway, N. A., Glickman, M., King, R. W., and Finley, D. (2007) *Nat. Struct. Mol. Biol.* **14**, 1180–1188
- Book, A. J., Yang, P., Scalf, M., Smith, L. M., and Vierstra, R. D. (2005) *Plant Physiol.* **138**, 1046–1057
- Rosenzweig, R., and Glickman, M. H. (2008) *Biochem. Soc. Trans.* **36**, 807–812
- Paz-Avram, T., Yahalom, A., and Chamovitz, D. A. (2008) *Plant Signal. Behav.* **3**, 409–411
- Peng, Z. H., Shen, Y. P., Feng, S. H., Wang, X. P., Chitteti, B. N., Vierstra, R. D., and Deng, X.-W. (2003) *Curr. Biol.* **13**, R504–R505
- Sadre-Bazzaz, K., Whitby, F. G., Robinson, H., Formosa, T., and Hill, C. P. (2010) *Mol. Cell* **37**, 728–735
- Smalle, J., Kurepa, J., Yang, P., Emborg, T. J., Babiychuk, E., Kushnir, S., and Vierstra, R. D. (2003) *Plant Cell* **15**, 965–980
- Gomes, A. V., Zong, C., Edmondson, R. D., Li, X., Stefani, E., Zhang, J., Jones, R. C., Thyparambil, S., Wang, G. W., Qiao, X., Bardag-Gorce, F., and Ping, P. (2006) *Circ. Res.* **99**, 362–371
- Wang, X., Chen, C. F., Baker, P. R., Chen, P. L., Kaiser, P., and Huang, L. (2007) *Biochemistry* **46**, 3553–3565

56. Satoh, K., Sasajima, H., Nyomura, K. I., Yokosawa, H., and Sawada, H. (2001) *Biochemistry* **40**, 314–319
57. Rivett, A. J., Bose, S., Brooks, P., and Broadfoot, K. I. (2001) *Biochimie* **83**, 363–366
58. Kimura, Y., Saeki, Y., Yokosawa, H., Polevoda, B., Sherman, F., and Hirano, H. (2003) *Arch. Biochem. Biophys.* **409**, 341–348
59. Kimura, Y., Takaoka, M., Tanaka, S., Sassa, H., Tanaka, K., Polevoda, B., Sherman, F., and Hirano, H. (2000) *J. Biol. Chem.* **275**, 4635–4639
60. Boisson, B., Giglione, C., and Meinnel, T. (2003) *J. Biol. Chem.* **278**, 43418–43429
61. Chen, P., and Hochstrasser, M. (1996) *Cell* **86**, 961–972
62. Ermolaeva, M. D., Wu, M., Eisen, J. A., and Salzberg, S. L. (2003) *Plant Mol. Biol.* **51**, 859–866
63. Kim, H. J., Oh, S. A., Brownfield, L., Hong, S. H., Ryu, H., Hwang, I., Twell, D., and Nam, H. G. (2008) *Nature* **455**, 1134–1137
64. Kloetzel, P. M. (2001) *Nat. Rev. Mol. Cell Biol.* **2**, 179–187
65. Kurepa, J., Toh-E, A., and Smalle, J. A. (2008) *Plant J.* **53**, 102–114
66. Elsasser, S., Gali, R. R., Schwickart, M., Larsen, C. N., Leggett, D. S., Müller, B., Feng, M. T., Tübing, F., Dittmar, G. A., and Finley, D. (2002) *Nat. Cell Biol.* **4**, 725–730
67. Pick, E., Hofmann, K., and Glickman, M. H. (2009) *Mol. Cell* **35**, 260–264
68. Lee, K. K., Florens, L., Swanson, S. K., Washburn, M. P., and Workman, J. L. (2005) *Mol. Cell Biol.* **25**, 1173–1182
69. Kim, Y. K., Kim, Y. S., Yoo, K. J., Lee, H. J., Lee, D. R., Yeo, C. Y., and Baek, K. H. (2007) *Gene Expr. Patterns* **7**, 143–148