Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach

Keren Lasker[a,b,1], Friedrich Förster[c,1], Stefan Bohn[c], Thomas Walzthoeni[d,e], Elizabeth Villa[c], Pia Unverdorben[c], Florian Beck[c], Ruedi Aebersold[d,f], Andrej Sali[a,2], and Wolfgang Baumeister[c,2]

[a]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute of Quantitative Biosciences, 1700 4th Street, University of California, San Francisco, CA 94158; [b]Blavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel; [c]Department of Molecular Structural Biology, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany; [d]Department of Biology, Institute of Molecular Systems Biology, Eidgenössische, Technische Hochschule, 8093 Zürich, Switzerland; [e]PhD Program in Molecular Life Sciences, University of Zurich/ETH Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; and [f]Faculty of Science, University of Zürich, 8093 Zürich, Switzerland

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2010.

Contributed by Wolfgang Baumeister, December 14, 2011 (sent for review December 9, 2011)

The 26S proteasome is at the executive end of the ubiquitin-proteasome pathway for the controlled degradation of intracellular proteins. While the structure of its 20S core particle (CP) has been determined by X-ray crystallography, the structure of the 19S regulatory particle (RP), which recruits substrates, unfolds them, and translocates them to the CP for degradation, has remained elusive. Here, we describe the molecular architecture of the 26S holocomplex determined by an integrative approach based on data from cryoelectron microscopy, X-ray crystallography, residue-specific chemical cross-linking, and several proteomics techniques. The "lid" of the RP (consisting of Rpn3/5/6/7/8/9/11/12) is organized in a modular fashion. Rpn3/5/6/7/9/12 form a horseshoe-shaped heterohexamer, which connects to the CP and roofs the AAA-ATPase module, positioning the Rpn8/Rpn11 heterodimer close to its mouth. Rpn2 is rigid, supporting the lid, while Rpn1 is conformationally variable, positioned at the periphery of the ATPase ring. The ubiquitin receptors Rpn10 and Rpn13 are located in the distal part of the RP, indicating that they were recruited to the complex late in its evolution. The modular structure of the 26S proteasome provides insights into the sequence of events prior to the degradation of ubiquitylated substrates.

coiled coils | mass spectrometry | proteasome-COP9-eIF3 domain | proteasome-cyclosome repeats

In eukaryotes, the ubiquitin-proteasome pathway (UPP) is essential for proteostasis: Misfolded proteins or otherwise defective proteins as well as short-lived regulatory proteins are eliminated by degradation (1). The UPP regulates many fundamental cellular processes, such as protein quality control, DNA repair, and signal transduction (for review see ref. 2). The 26S proteasome is the most downstream element of the UPP, executing the degradation of polyubiquitylated substrates (3–5). It consists of the barrel-shaped core particle (CP; approximately 700 kDa), which sequesters the proteolytically active site in its central cavity, and the regulatory particle (RP; approximately 900 kDa), which is attached at either one or both ends of the CP and prepares substrates for degradation (6).

The RP consists of 19 different canonical subunits, including six regulatory particle AAA-ATPase subunits (Rpt1-6) and 13 regulatory particle non-ATPase subunits (Rpn1-3, Rpn5-13, and Rpn15). The integral ubiquitin (Ub) receptors Rpn10 and Rpn13 recognize polyubiquitylated substrates (7–9). Alternatively, polyubiquitylated substrates can be recruited by shuttling Ub-receptors (Dsk2, Rad23, Ddi2), which bind to substrates with their Ub-associated domain, and to Rpn1, Rpn10, or Rpn13 at their Ub-like domain (5). The metalloprotease Rpn11 deubiquitylates substrates prior to their degradation (10, 11). The functions of the other Rpn subunits remain to be established. The AAA-ATPases form a hexameric ring that unfolds substrates, opens the gate to the CP, and eventually translocates the substrates to the CP.

Electron microscopy (EM) (12) and X-ray crystallography studies of the CP revealed an evolutionary highly conserved structure, namely a stack of four seven-membered rings (α1-7; β1-7; β1-7; α1-7) (13, 14). The chamber formed by the β-subunits houses the proteolytic sites and the outer α-rings constitute two antechambers (15) serving as holding compartments (16). Most of our structural knowledge on the AAA-ATPases so far stems from the archaeal homolog of the RP, the proteasome-activating nucleotidase (PAN) (17, 18). The six AAA-ATPase subunits assemble into two stacked rings: The ring formed by the AAA-domains (AAA-ring) is pseudo sixfold symmetrical and positioned proximal to the CP, whereas the distal N-terminal domains form a pseudo threefold symmetric ring (trimer of dimers) with coiled-coil pairs protruding from it (N ring). The N ring unfolds substrates whereas the AAA ring is believed to actively translocate substrates into the CP in an ATP-dependent manner (17, 19). The topological order of the AAA-ATPase subunits is Rpt1/Rpt2/Rpt6/Rpt3/Rpt4/Rpt5, as determined based on protein–protein interactions (20) and engineered disulfide cross-links (21). Integration of a cryoelectron microscopy (cryo-EM) map at 9.1-Å resolution, protein–protein interactions, and site-specific intermolecular cross-links revealed the relative positions of the AAA-ATPase ring and the CP (20, 22), which was subsequently corroborated by engineered disulfide cross-links between the AAA-ATPase C termini and residues in the pockets between the α-subunits of the CP (23). In contrast, the spatial configuration of the non-ATPase subunits and their role in the preparation of substrates for degradation remains largely unknown.

The RP has been refractory to any single structure determination method, presumably due to sample heterogeneity caused by the RP's tendency to disassociate into heterogeneous subcomplexes during purification and concentration as well as the presence of proteasome interacting proteins (PIPs). Moreover, some RP subunits may also exhibit a significant degree of structural variability. As a consequence, X-ray crystallographic analysis of

Author contributions: R.A., A.S., and W.B. designed research; K.L., F.F., S.B., T.W., E.V., and P.U. performed research; K.L., F.F., S.B., T.W., E.V., P.U., F.B., and A.S. analyzed data; and K.L., F.F., R.A., A.S., and W.B. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The single particle reconstruction has been deposited in the Electron Microscopy Data Bank, http://www.ebi.ac.uk/pdbe/emdb/ (accession code 10440).

[1]K.L. and F.F. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: sali@salilab.org or baumeist@biochem.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120559109/-/DCSupplemental.

the entire RP has not yet been accomplished, although the atomic structures of a few subunit fragments have been determined (9, 24–26). To overcome limitations in sample preparation and individual structure-determination techniques, we have applied an integrative structure determination approach in which a variety of different datasets were computationally integrated into a more accurate and higher-resolution structural model than is possible with any of the individual datasets (27–29). In particular, we have determined the structure of the complete 26S proteasome from *Schizosaccharomyces pombe*, based on a new 8.4-Å cryo-EM map, residue-specific intersubunit cross-links, crystallographic structures of RP subunit homologs, as well as previously published protein–protein interaction data from a variety of species. The modular architecture of the RP provides unique insights into the sequence of events prior to degradation.

Results and Discussion

8.4-Å Resolution Cryo-EM Map of the 26S Proteasome. We have used cryo-EM to structurally characterize the 26S proteasome holocomplex from *Schizosaccharomyces pombe* (Fig. 1*A*). Notably, we have approximately doubled the number of particles previously used for a single-particle reconstruction (22), to 375,000. This increase allowed us to improve the resolution from 9.1 to 8.4 Å (Fig. S1). The resolution of the cryo-EM map varies locally (Fig. 1*B*); the CP is better resolved than the RP, which exhibits "hot spots" of structural variability. From the cryo-EM map we segmented the CP and the AAA-ATPase modules based on the CP and PAN crystal structures (20, 22) (Fig. 1*A*). The remaining cryo-EM density corresponds to the Rpn subunits. Numerous helical elements, many of which resemble bihelical repeats, are clearly visible in the 8.4-Å resolution map.

Subunit Interactions Revealed by Chemical Cross-Linking/Mass Spectrometry (XL/MS). To obtain information on protein–protein proximities at the peptide level, we chemically cross-linked the purified *S. pombe* 26S proteasomes using the lysine reactive cross-linker disuccinimidyl suberate (DSS), subjected the trypsin-digested samples to liquid chromatography-tandem mass spectrometry (LC-MS/MS), and identified the cross-linked peptides from MS/MS spectra at a false-discovery rate (FDR) of less than 5% (*Materials and Methods*) (30). Thirty-two cross-links were obtained within the AAA-ATPase hexamer and four between the CP and the AAA-ATPases, involving nine and three different subunit pairs, respectively (Fig. 2 and Table S1); all of these cross-links are consistent with our previously determined subunit organization (20). An additional 15 intra-Rpn and eight Rpt/Rpn cross-links involved

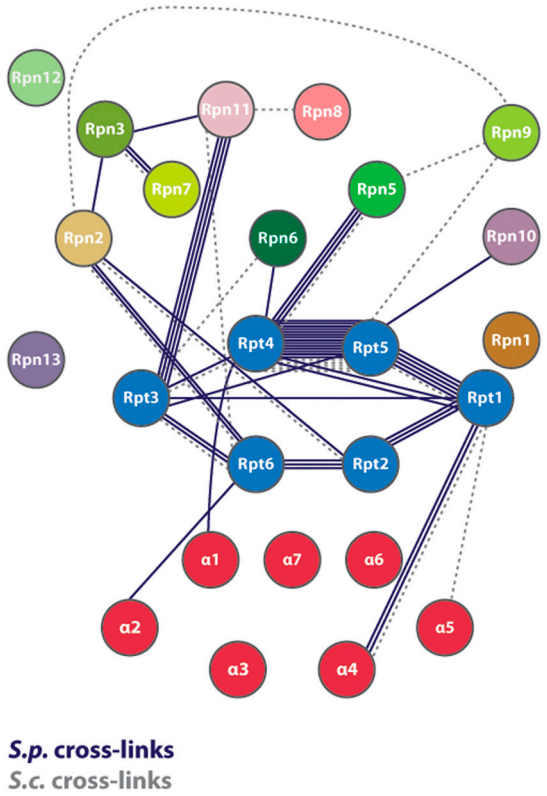

Fig. 2. Chemical cross-links for the *S. pombe* and *S. cerevisiae* 26S proteasomes. Fifty-five (21) pairs of cross-linked lysines from the *S. pombe* (*S. cerevisiae*) 26S proteasome subunits are shown as an interaction network (Table S1). Multiple edges between a pair of subunits indicate multiple cross-linked lysine pairs.

seven out of the 13 Rpn subunits and 15 different subunit pairs (Fig. 2 and Table S1). Notably, these cross-links indicate proximity of the lysine residues in the N-terminal segments of the non-ATPases Rpn5 and Rpn6 to the AAA domains of Rpt3 and Rpt4.

While we used the data from *S. pombe* 26S proteasome for integrative structure determination (below), we also performed cross-linking experiments with the 26S proteasome from *Saccharomyces cerevisiae* for validation. Not surprisingly, four of the 11 *S. cerevisiae* cross-links apply to the same Rpn/Rpn and Rpn/Rpt pairs as in the *S. pombe* dataset, indicating their accuracy. Never-

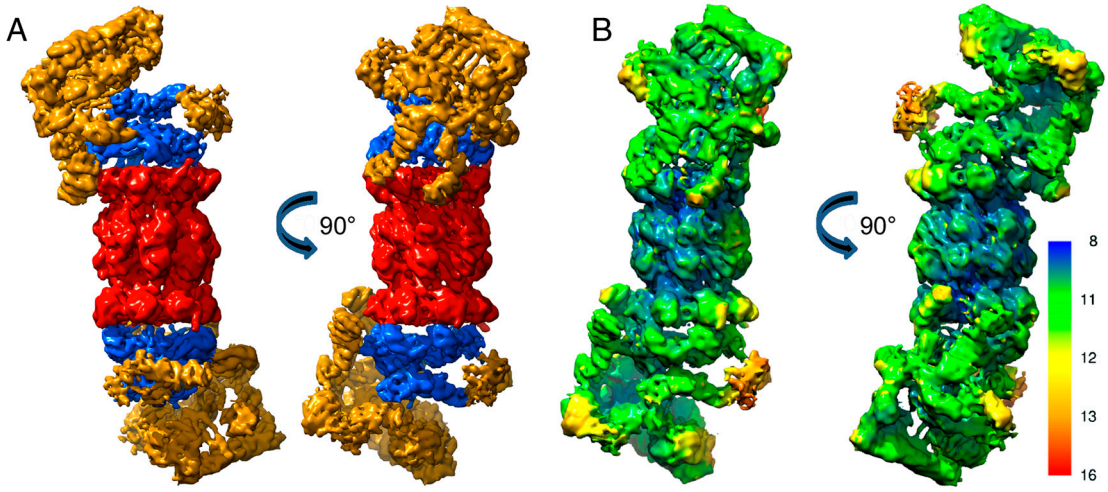

Fig. 1. Cryo-EM map of the *S. pombe* 26S proteasome. (*A*) The single-particle cryo-EM density map of the 26S proteasome from *S. pombe* at 8.4-Å resolution is shown in two views, related by a 90° rotation around the pseudo-sevenfold axis of the CP (CP: red; AAA-ATPase hexamer: blue; Rpn subunits: gold). (*B*) The isosurface of the cryo-EM map is colored according to the local resolution in Å, as specified in the color bar.

theless, some cross-links observed in the *S. pombe* dataset were not observed in *S. cerevisiae* and vice versa. Thus, complementary information can be obtained by cross-linking protein complexes of related organisms. The differences in cross-link coverage are attributed to the differences in the lysine residue locations between the two species (only two of the 11 Rpn/Rpt and Rpn/Rpn cross-links occurred at corresponding lysine pairs) and to technical limitations of XL/MS.

Integrative Structure Determination of the RP Architecture. To determine the positions, orientations, and conformations of the Rpn subunits and the AAA-ATPase hexamer, we applied an integrative approach implemented in our open source Integrative Modeling Platform (IMP) package (http://salilab.org/imp/) (27, 28, 31). The process consists of four stages: gathering information about the structure of the 26S proteasome, choosing how to represent the system and how to translate the information into spatial restraints, calculating an ensemble of structures that satisfy these restraints, and analyzing the ensemble (Fig. 3). The approach has been previously used for determining the positions of 456 subunits in the eightfold symmetric ring of the yeast nuclear pore complex (31, 32). However, the 26S proteasome has presented new challenges—namely, the asymmetry of the assembly and the interpretation of the structure at the amino acid level. These challenges required integration of new types of data, a multiscale model representation including the atomic level, as well as a more powerful structural sampling scheme.

Stage 1: Gathering Information. Five different types of information were used for structure determination: (*i*) an 8.4-Å cryo-EM map of the RP from *S. pombe*; (*ii*) 12 residue-specific cross-links between the AAA-ATPase and Rpn subunits as well as three cross-links between the Rpn subunits from *S. pombe* (Table S1); (*iii*) densities of Rpn10 and Rpn13 in the 26S holocomplex as revealed by cryo-EM of the *S. cerevisiae* 26S proteasomes from ΔRpn10 and ΔRpn13 deletion strains (33); (*iv*) 47 direct interactions and proximities between the RP subunits determined by various proteomics studies, including two-hybrid assays, chemical cross-linking, and immunoprecipitations (Table S2); as well as (*v*) atomic structures of the individual Rpn subunits and their homologs (Fig. S2 and Table S3).

Stage 2: System Representation and Translation of Information into Spatial Restraints. Two representations of a sequence segment, subunit, or a subcomplex were used, depending on the data available, spatial restraints applied, and sampling performed (Table S4). The first, coarse-grained representation depicted each approximately 50-residue fragment of the segment by a bead. If the atomic structure of a subunit or subcomplex (i.e., the AAA-ATPase hexamer) was predicted by comparative modeling, the beads were fixed to mimic the shape of the atomic model; otherwise, the string of beads was allowed to flex. The second, atomic representation specified atomic coordinates for segments with available comparative models.

The spatial restraints encoded the different types of information gathered (Table S5). First, a measure of cross-correlation between the cryo-EM map and the atoms in an RP model restrained the shape of the model. Second, upper distance bounds enforced proximity between pairs of cross-linked atoms or subunits, depending on the representation used. Third, the "connectivity" restraints (31) enforced direct protein interactions and proximities among the subunit sets in the RP subcomplexes (Table S2). In addition, we applied a radius-of-gyration restraint to each flexible subunit in the coarse-grained representation (31), an excluded-volume restraint to each bead and/or atom pair, and a geometric-complementarity restraint for each pair of interacting subunits in the atomic representation.

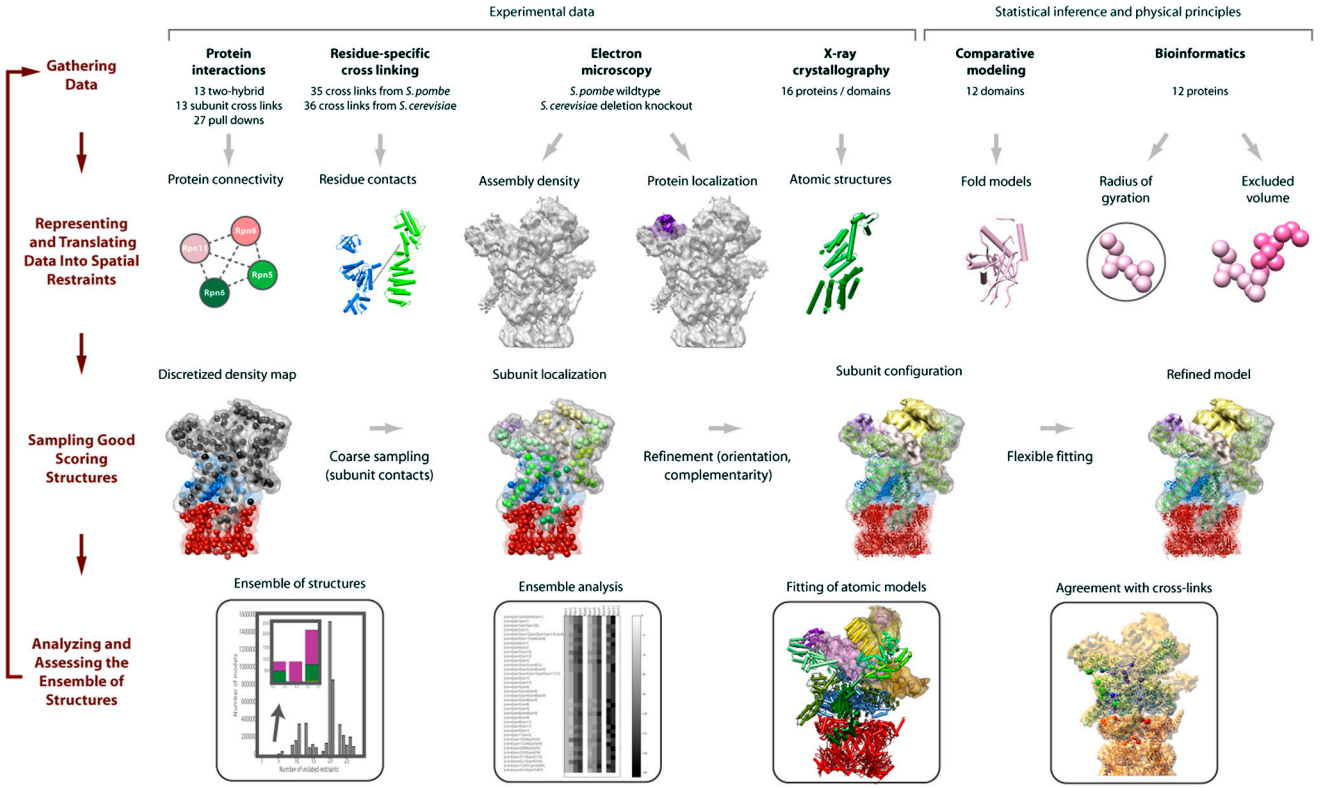

Fig. 3. Integrative structure determination of the RP. First, structural data and information are generated by various experiments and computational methods, respectively. Second, the AAA-ATPase hexamer and the Rpn subunits are each represented as a string of beads or by atoms, and the data are translated into spatial restraints on their arrangement. Third, an ensemble of structures that satisfy the data are obtained by inferential sampling. Fourth, the ensemble is clustered into distinct subsets of structures, and analyzed in terms of geometry and accuracy.

Stage 3: Ensemble Sampling. Models of the RP that satisfy the data as well as possible were obtained in three steps. In the first, subunit localization step, best-scoring models were enumerated in the coarse-grained representation by aligning the RP interaction network implied by the residue-specific cross-links (Table S1) and subunit interactions (Table S2) into the cryo-EM density map. The sampling started by discretizing the density map into a graph of 238 nodes, with graph nodes corresponding to the density map regions and graph edges connecting the neighboring regions. Possible localizations for each subunit were then found efficiently by applying a "message passing" algorithm to the graph (34). The sampling explored only localizations that were consistent with the subunit shapes and localization data. In particular, the positions of the Rpn10 and Rpn13 beads in the RP density were constrained by the corresponding difference cryo-EM maps. Similarly, the positions of the AAA-ATPases hexamer (20) and Rpn6 (26) were fixed based on previous fitting of their atomic models into a cryo-EM map. The number of candidate localizations for each nonlocalized subunit ranged from 10^3 for the small Rpn subunits to 10^5 for the large Rpn1 and Rpn2 subunits (Table S4). Approximately $5 \cdot 10^5$ best-scoring coarse-grained RP models that significantly violated at most five restraints were finally found by enumerating combinations of subunit localizations, again using message passing.

In the second, subunit fitting step, multiple comparative models were computed for each of the *S. pombe* Rpn subunits (*Materials and Methods* and Fig. S2 and Table S3). Alternative combinations of comparative models of the PCI subunits were then simultaneously fitted into the region of the cryo-EM density map identified uniquely in the previous subunit localization step using MultiFit (34) (*Materials and Methods* and Fig. S3). Finally, additional high-confidence models of Rpn8, Rpn10, and Rpn2 were fitted to the map (Fig. S4) using MOLMATCH (35).

In the third, subunit refinement step, the atomic coordinates of the PCI subunits were refined by flexible fitting using molecular dynamics flexible fitting (MDFF) (36).

Stage 4: Assessment of Data and Structure. We analyzed the ensemble of approximately $5 \cdot 10^5$ best-scoring RP coarse-grained models in terms of subunit positions, contacts, and configuration. The models were clustered according to the subunit centroid rmsd, revealing three large clusters of models with an average centroid rmsd within each cluster of only 5 Å (Fig. 4 *A* and *B*). The largest variability in the ensemble corresponds to a swap between Rpn8 and Rpn11 as well as between Rpn7 and Rpn3 (Figs. 4 *C* and *D*). Random subsets of 10% of the best-scoring models share a similar clustering pattern, confirming that our sampling procedure has indeed enumerated the best-scoring models, given the data and the sampling resolution.

The confidence in the accuracy of the ensemble is based on three considerations (27, 31). First, the ensemble models satisfy most of the independently determined data points. All good scoring models violate the same small subset of restraints, including the residue-specific cross-links Rpn10:30/Rpt5:54 and Rpn6:272/Rpt4:164, the Rpn10-Rpn1 interaction (37, 38), and the Rpt1/2, Rpn2/8/12/13 subcomplex (39). The first three violations could be explained by conformational variability (see below) and the fourth one by an error or incompleteness of high-throughput detection. Second, structural similarity among the models in the ensemble, as described above, is sufficient to determine the posi-

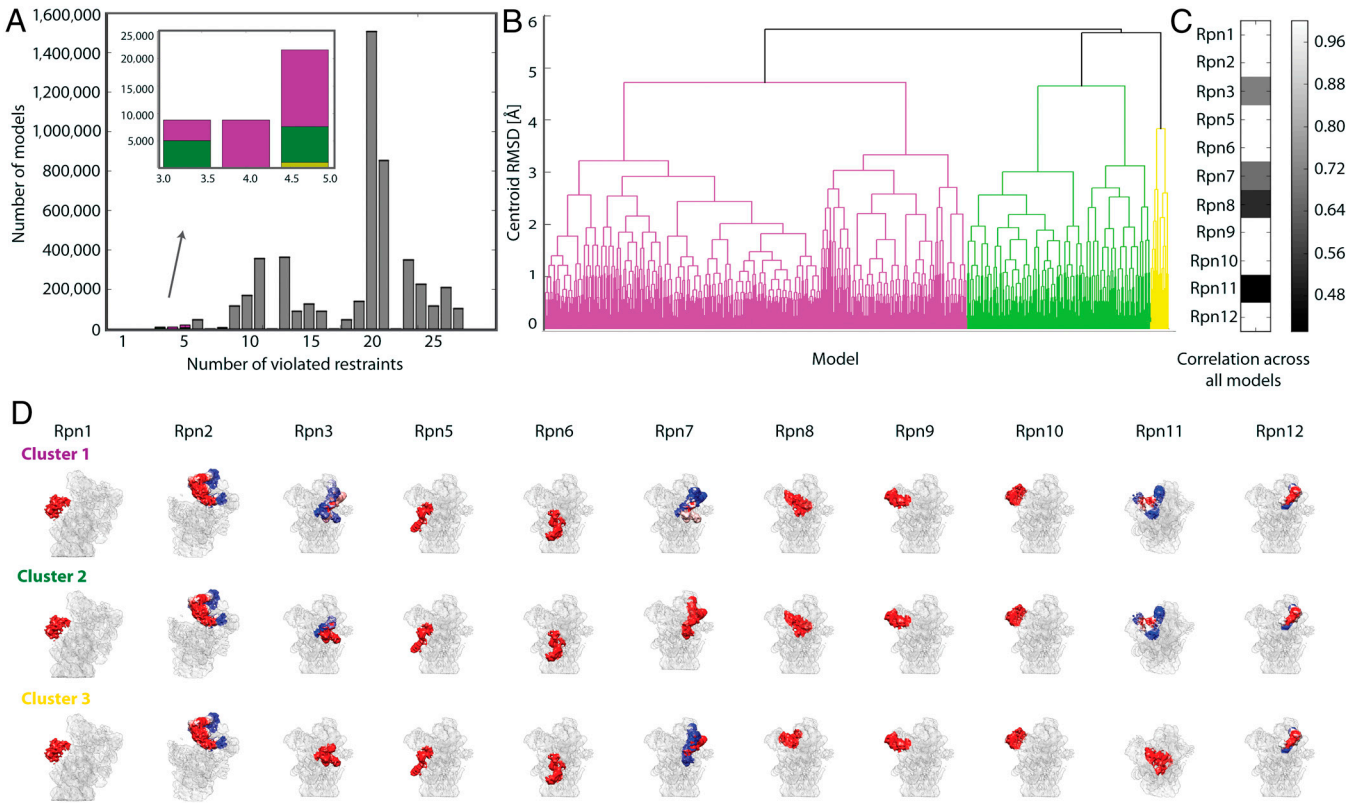

Fig. 4. Analysis of the ensemble of 26S proteasome structures. (*A*) Distribution of scores for the ensemble of models. $5 \cdot 10^5$ of the nearly 250 million sampled models violated at most five of the 98 cross-linking and proteomics restraints. (*B*) Clustering of the top scoring models that violate at most five cross-linking and proteomics restraints. The models were clustered by their centroid rmsd into three main clusters with the maximum rmsd from the cluster centroid of 5 Å. (*C*) Conservation of subunit positions in the ensemble. For each subunit, the average cross-correlation coefficient between pairs of subunit clusters is displayed. The plot shows localization variability for Rpn3, Rpn7, Rpn8, and Rpn11. (*D*) Localization probabilities for each subunit in the cluster displayed as a heat map. Each voxel of the density map is colored by its localization probability, defined as the percentage of cluster models that localized a subunit in the voxel. A gradient color from blue to red indicates that the subunit was localized to a voxel in 5% to 100% of the models; a gray color indicates 0%.

tions of most subunits and orientations for subunits with atomic models. Third, the ensemble is consistent with the structural data not used to compute it, including 21 residue-specific intersubunit cross-links from *S. cerevisiae*, the PCI domain contacts implied by the Rpn6 crystal contacts (26), the high-confidence fits of the atomic models of the leucine rich repeat (LRR) region of Rpn2 (40) (Fig. S4), the peripheral position of Rpn12 from the whole complex mass spectrometry (41), the Rpn9 interaction with the Ub receptor Rpn10 from a two-hybrid assay (42), and the subcomplexes formed during lid assembly (43). Moreover, the ensemble is consistent with >95% of the XL/MS data at the residue level, although cross-links were imposed only at the subunit-level (Fig. S5). Finally, even after omitting one restraint at a time, the resulting jackknife ensembles remain similar to the original ensemble (Fig. S6).

AAA-ATPase Hexamer. To analyze the AAA-ATPase hexamer in more detail, its comparative model based on the PAN structures (22) was subjected to flexible fitting into the EM map (36). The refined model reveals substantial deviations from the approximate sixfold rotational symmetry of the PAN AAA ring (Fig. 5*A*). In particular, the Rpt3, Rpt6, and Rpt4 subunits vary from those in the symmetrical starting model. Moreover, the principal axes of the N ring and the AAA ring are not in register. Thus, the evolutionary diversification of the Rpt subunit functions (44) has also resulted in a major divergence of their structures. The cryo-EM map allowed us to resolve the C termini of the Rpt2 and Rpt3 subunits, which are absent in the PAN crystal structures (Fig. 5*B*). These two C termini are located in the pockets between the α3/α4 and α1/α2 CP subunits, respectively. Rpt2 and Rpt3 both contain the C-terminal HbYX motif that has been shown to induce gate opening of the CP (45). We also observed density in the pocket formed by the α5/α6 pocket, which likely corresponds to the C terminus of Rpt5, the third Rpt subunit containing HbYX motif. However, this density is weaker than that observed in the α3/α4 and α1/α2 pockets. A recent cross-linking study, in which cysteines were introduced into the AAA-ATPase C termini, was consistent with our observations (23). Specifically, Rpt2 and Rpt3 were found to localize in the pockets formed by the CP α3/α4 and α1/α2 interfaces, respectively, whereas Rpt5 was not confined to a specific pocket and was found to occupy both α5/α6 and α6/α7 pockets.

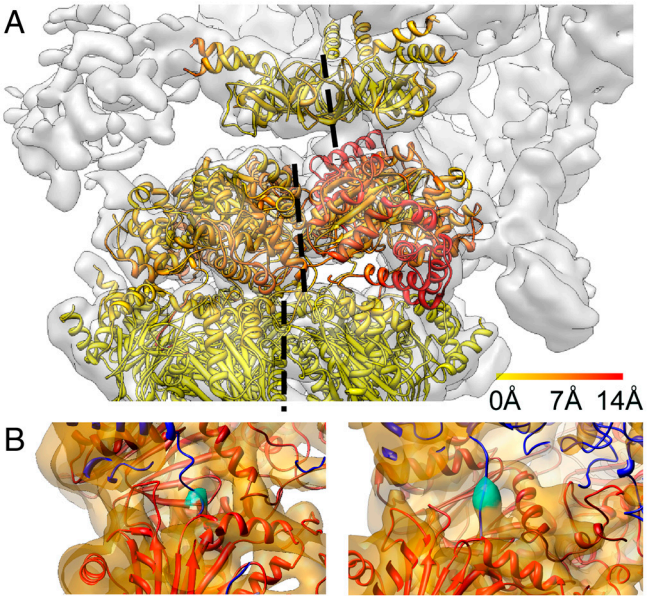

Fig. 5. Structure of the AAA-ATPase hexamer. (*A*) Atomic representation of the CP and AAA-ATPase hexamer after flexible fitting into the cryo-EM map, colored by their C^{α} atom rmsd compared to the initial comparative models based on the *S. cerevisiae* CP (13) and the AAA/N-rings from PAN (18). (*B*) Near the C termini of Rpt2 (left) and Rpt3 (right), we observe density in the pockets formed by the α3/α4 and α1/α2 interfaces, respectively (cyan).

Modular Architecture of Non-ATPases. We determined the configuration of all Rpn subunits relative to the AAA-ATPase and CP. Based on sequence similarity, the Rpns can be divided into four categories (46): (*i*) the proteasome cyclosome (PC) repeat containing subunits Rpn1 and Rpn2, (*ii*) the MPN (Mpr1 and Pad1 in the N terminus) domain containing subunits Rpn8 and Rpn11, (*iii*) the proteasome-COP9-eIf3 (PCI) domain containing subunits Rpn3/5/6/7/9/12, and (*iv*) the Ub receptors Rpn10 and Rpn13. These evolutionary relationships are reflected in the modular structure of the RP.

PC-Repeat-Containing Subunits. The hallmarks of the Rpn1 and Rpn2 sequences are nine-repeat sequences (PC repeats), which also occur in the cyclosome/anaphase-promoting complex subunits (47). The PC repeats are predicted to resemble bihelical LRRs, which assemble into a torus (40, 48). In our structure, we find Rpn1 at the periphery of the Rpt1/2 dimer (Fig. 6*A*). The density corresponding to Rpn1 is highly variable (Fig. 1B), as already observed (49), which makes it impossible to discern secondary structure elements. This variability may be due to Rpn1's function as a recruitment factor for shuttling Ub receptors and the DUB Ubp6, which dock to the LRRs of Rpn1 (50, 51).

In contrast to Rpn1, Rpn2 is structurally well defined. With its domain proximal to the CP, it contacts the Rpt3/6 dimer and projects to the distal part of the RP. Short helices are discernible at many positions in the Rpn2 density, suggesting that Rpn2 consists largely of helical repeats. Interestingly, a segment of Rpn2 has a toroidal shape, as predicted for the PC-repeat region (40, 48). Indeed, the proposed atomic model (40) matches the cryo-EM density map very well in this region (Fig. S4). Rpn2 contacts the Rpn8/11 dimer at its torus-shaped region and it interacts with Rpn12 and Rpn3 at its distal end. The Ub receptor Rpn13 also binds to the distal, probably C-terminal domain of Rpn2 (9).

MPN-Domain-Containing Subunits. Rpn8 and Rpn11 both contain an N-terminal MPN domain. The MPN domain of Rpn11, in contrast to that of its paralog Rpn8, contains a zinc-binding JAMM (JAB1/MPN/Mov34 metalloenzyme) motif, which is responsible for its deubiquitylation activity (24, 52, 53). Rpn8 and Rpn11 form a heterodimer, which is positioned above the opening of the AAA-ATPase ring and incorporated into the scaffold provided by the PCI-domain-containing subunits (Fig. 6). While the heterodimer was localized with high confidence, the individual Rpn8 and Rpn11 positions remain ambiguous (only the MPN domain of Rpn8 can be positioned accurately) (Fig. S4).

PCI-Domain-Containing Subunits. We recently solved a protein structure containing a proteasomal PCI-domain (26): Rpn6 consists of several bihelical repeats that assemble into a solenoid fold followed by a PCI domain (Fig. S4). Structure prediction suggests that all PCI-domain-containing subunits have the same fold (Fig. S2). Protein–protein interaction data show that the PCI-containing subunits colocalize in the same region of the RP (46). Some of these interactions have been narrowed down to their PCI domains (26, 54). Our structure reveals that the six PCI subunits are arranged in a horseshoe-like fashion forming a "roof" covering a large part of the AAA-ATPase heterohexamer (involving Rpt3, Rpt6, and Rpt4) (Fig. 6). The subunit order in the horseshoe is Rpn9/Rpn5/Rpn6/Rpn7/Rpn3/Rpn12. The PCI-heterohexamer embraces the Rpn8/11 heterodimer and contacts Rpn10 (via Rpn9) and Rpn2 (via Rpn12) at its ends. The scaffold of the horseshoe is made of the interacting PCI domains, while the α-solenoid domains project away from the horseshoe (Fig. 6*D*). Interestingly, Rpn6 monomers contact each other through their PCI domains in the Rpn6 crystal and form a screw (26), which

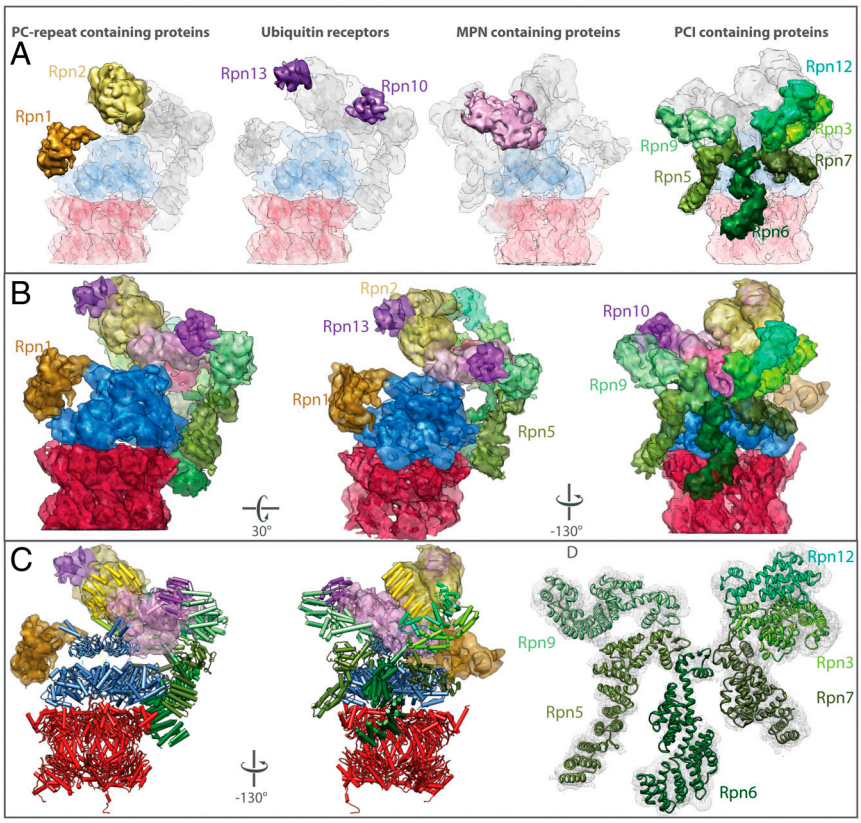

Fig. 6. Modular architecture of the 26S proteasome. (*A*) From left to right: segmented densities of the PC-repeat containing proteins, Ub receptors (Rpn13 is taken from the *S. cerevisiae* 26S proteasome because it binds substoichmetrically in *S. pombe*) (33), the MPN-domain-containing proteins, and the PCI-domain-containing subunits. (*B*) The segmented 26S proteasome density in three different views. (*C*) Hybrid representation of the 26S proteasome, indicating fitted available comparative models for the CP (red), the AAA-ATPase heterohexamer (blue), the PCI heterohexamer (green), Rpn8 MPN domain (pink), Rpn2 LRR domain (yellow), and Rpn10 VWA domain (purple). (*D*) Close-up view of the heterohexamer formed by the PCI domains.

resembles the horseshoe observed for the PCI-heterohexamer in the RP. The solenoid N termini of the PCI-domain-containing subunits all form right-handed superhelices, but their respective curvatures vary substantially.

Ub Receptors. Rpn10 and Rpn13 are structurally unrelated and recognize Ub through different binding motifs (8). From their peripheral localization at the distal end of the 26S proteasome, we conclude that both receptors were added late in the evolution of the RP. Rpn13 interacts solely with Rpn2, which is consistent with numerous interaction studies (9, 55–57). We anticipated that the cryo-EM map could only resolve the C-terminal von Willenbrand factor A (VWA) domain of Rpn10, due to the structural variability of the N-terminal ubiquitin interacting motif (UIM) (58), which is consistent with the fit for Rpn10 (Fig. S4*D*). Rpn10 interacts with Rpn9 and the Rpn8/11 heterodimer. We do not observe the Rpn10–Rpn1 interaction, which was implied by immunoprecipitation (38, 59); nevertheless, this interaction may occur transiently. The Rpn10 UIM is flexible, and in the cryo-EM map we observe Rpn1 to be variable. Hence, Rpn1 and the Rpn10 UIM may bridge the distance of approximately 70 Å between Rpn1 and the Rpn10 VWA domain temporarily, which is not resolved in our map.

Structure and Function of the Lid Subcomplex. The subcomplex formed by the MPN- and PCI-domain-containing subunits was termed the "lid" (60), indicating that it was positioned distally from the CP. However, it turns out that the lid instead flanks the side of the AAA-ATPase hexamer and even contacts the CP directly; the N termini of Rpn6 and to a lesser extent Rpn5 physically interact with the CP. The only currently known function of the lid is a deubiquitylating activity. Rpn11 is only active within the assembled RP and its deubiquitylating activity is ATP-dependent (10, 11). We find the Rpn8/Rpn11 heterodimer close to the AAA-ATPase hexamer, which may explain the ATP-dependence of the Rpn11 function. The positioning of Rpn8/Rpn11 suggests that Rpt3 and Rpt6 may be critical for activation of Rpn11. Nevertheless, higher resolution studies are required to precisely identify the mechanism by which the AAA-ATPases activate Rpn11 and to elucidate the role of Rpn8.

The AAA-ATPase heterohexamer is thought to undergo large-scale motions during its functional cycle (61), probably similar to the ClpX hexamer (62). One function of the hexamer formed by the PCI-containing subunits seems to be that of a scaffold, which positions Rpn11 near the mouth of the AAA-ATPase. Pivotal for this scaffolding function is Rpn6, which maintains the contact to the CP through its solenoid N terminus. A further role of the PCI submodule at the top of the AAA-ATPase module could be that of a shield: The AAA-ATPases are expected to have a propensity to bind to secondary degradation signals (degrons) of substrates (see below) and their shielding could help to minimize uncontrolled degradation.

It is unlikely that a complex structure such as the PCI horseshoe has evolved to solely function as a scaffold or a shield. For example, the 71-residue subunit Rpn15 (Sem1), which is associated with Rpn3 and Rpn7, would not be required if the PCI subunits did not have further functions (63). The PCI domain contains a winged-helix motif, which typically binds nucleotides (64). Interestingly, when the winged-helix motif of Rpn6 is superposed on that of the transcription factor E2F-DP (PDB ID code 1CF7), the bound DNA molecule would not clash with the surrounding proteasome subunits (Fig. S7). DNA association with

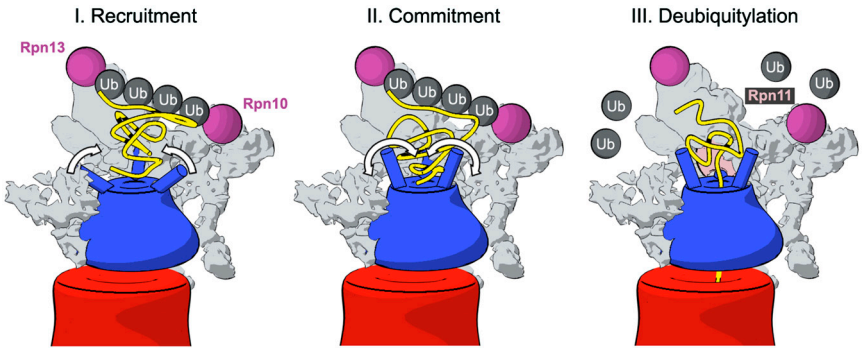

Fig. 7. Mechanistic model for the early steps of protein degradation by the 26S proteasome. Polyubiquitylated substrates first bind to the Ub receptors Rpn10 and Rpn13 in concert (33) (*I*). During the subsequent commitment step (*II*), the substrates are more tightly bound by the AAA-ATPase coiled coils, which may perform swinging motion. The coiled coils further transfer the substrates to Rpn11, where they are deubiquitylated to enable Ub recycling (*III*).

the 26S proteasome could be required for its role in transcription (65), but no experimental evidence for recruitment of the 26S proteasomes to DNA or RNA exists to date. Interestingly, for subunit eIF3c of the lid paralog eIF3, RNA interaction of its PCI domain has been shown recently (66).

Model for the Initial Protein Degradation Steps. Based on the molecular architecture of the RP and biochemical findings, we put forward a model for the early steps of protein degradation by the 26S proteasome (Fig. 7). Genetic studies revealed a functional linkage between Rpn10 and Rpn13 (8). In NMR spectroscopy experiments, purified Rpn13 and Rpn10 were found to simultaneously bind to K-48 linked di-ubiquitin (67). Together, these findings suggest that Ub receptors may bind polyubiquitin tags "in concert" when recruiting substrates to the 26S proteasome (33). This hypothesis is consistent with the positions of Rpn10 and Rpn13 found in our structure. The distance between both subunits is approximately 90 Å, which could be bridged by a tetraubiquitin moiety (33). In fact, this relative placement of Ub receptors offers an explanation why polyubiquitin chains need to comprise at least four Ub subunits to function efficiently as a degradation signal (68, 69). Thus, we hypothesize that Rpn13 and Rpn10 both bind to the polyubiquitin tag of their respective substrates with a preference of Rpn13 for the proximal Ub moiety (67) (Fig. 7*I*).

Initial binding of polyubiquitylated substrates is followed by a "commitment" step (70); the substrate is bound more tightly to the proteasome via a loosely folded domain in an ATP-dependent manner. Rpn10 and Rpn13 are positioned above the coiled coils of the Rpt4/5 and Rpt1/2 dimers, respectively. The structure of the archaeal homolog PAN suggests that the N-terminal coiled coils are flexibly linked to the N ring of the ATPase (17). We envisage that they act as "swinging arms" sampling the space above the ATPase. Such swinging arms have been implicated in the transfer of substrates in multienzyme complexes to couple multistep reactions (71). Indeed, analysis of the space accessible to the coiled-coils pairs shows that the Rpt1/2 and Rpt4/5 coils can undergo substantial swinging motions without clashes with surrounding subunits (Fig. S8); our structural data indicate that the Rpt1/2 and Rpt4/5 coiled coils can rotate by up to approximately 40° around the hinge residues corresponding to Pro62 in PAN. This flexibility enables them to grab substrates bound to Rpn10 and Rpn13 via tetraubiquitin chains (33) and dangling into the space accessible to the swinging arms (Fig. 7*II*). It has already been shown that the PAN coiled coils bind substrates (72, 73), which suggests that the coiled coils have the capability to bind substrates with appropriate degrons (74).

Subsequently, the coiled coils expose the substrates to Rpn11 (Fig. 7*III*), which is located above the central opening of the AAA-ATPase. The substrates are deubiquitylated by Rpn11 before being unfolded in the upper AAA-ATPase cavity (N ring) and translocated to the CP where degradation occurs. The coiled-coil motions might trigger gate opening to the CP because the coiled coils are positioned above the Rpt subunits (Rpt2, Rpt3, and Rpt5) whose C termini open the gate (45).

Materials and Methods

Sample Preparation, Cryo-EM, and XL/MS. Sample preparation, cryo-EM single-particle analysis, and XL/MS were performed as described in ref. 22 and explained in more detail in *SI Methods*.

Integrative Structure Determination. In brief, multiple comparative models were built for each subunit, fitted into the cryo-EM map, and the best-scoring configurations were subjected to flexible fitting (*SI Methods*).

ACKNOWLEDGMENTS. We thank Julio Ortiz for preparing Fig. 7; Johannes Söding for help with HHpred; Daniel Russel, Ben Webb, and Charles Greenberg for help with IMP; Ursula Pieper for help with databases; as well as Eri Sakata, Stephan Nickell, Andreas Bracher, Ganesh Pathare, Istvan Nagy, Andreas Schweitzer, Alexander Leitner, Martin Beck, and Peter Cimermancic for fruitful discussions. This work was supported by grants from the European Union Seventh Framework Program PROSPECTS (Proteomics Specification in Space and Time Grant HEALTH-F4-2008-201648), the SFB 594 of the Deutsche Forschungsgemeinschaft (both to W.B.), the Human Frontier Science Project (Career Development Award to F.F.), a Marie Curie Postdoctoral Fellowship from the European Union (to E.V.), as well as National Institutes of Health Grants R01 GM54762, U54 RR022220, and R01 GM083960 (to A.S.). The research of K.L. was supported by continuous mentorship from Professor Haim J. Wolfson and a fellowship from the Clore Foundation PhD Scholars program.

1. Hershko A, Ciechanover A, Varshavsky A (2000) Basic Medical Research Award. The ubiquitin system. *Nat Med* 6:1073–1081.
2. Jentsch S, Haendler B (2009) *The Ubiquitin System in Health and Disease* (Springer, Heidelberg).
3. Voges D, Zwickl P, Baumeister W (1999) The 26S proteasome: A molecular machine designed for controlled proteolysis. *Annu Rev Biochem* 68:1015–1068.
4. Tanaka K (2009) The proteasome: Overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* 85:12–36.
5. Finley D (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* 78:477–513.
6. Peters JM, Cejka Z, Harris JR, Kleinschmidt JA, Baumeister W (1993) Structural features of the 26 S proteasome complex. *J Mol Biol* 234:932–937.
7. van Nocker S, et al. (1996) The multiubiquitin-chain-binding protein Mcb1 is a component of the 26S proteasome in *Saccharomyces cerevisiae* and plays a nonessential, substrate-specific role in protein turnover. *Mol Cell Biol* 16:6020–6028.
8. Husnjak K, et al. (2008) Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* 453(7194):481–488.
9. Schreiner P, et al. (2008) Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* 453:548–552.
10. Verma R, et al. (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* 298:611–615.
11. Yao T, Cohen RE (2002) A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* 419:403–407.
12. Puhler G, et al. (1992) Subunit stoichiometry and three-dimensional arrangement in proteasomes from *Thermoplasma acidophilum*. *EMBO J* 11:1607–1616.
13. Groll M, et al. (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* 386:463–471.
14. Lowe J, et al. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 A resolution. *Science* 268:533–539.

15. Baumeister W, Walz J, Zuhl F, Seemuller E (1998) The proteasome: Paradigm of a self-compartmentalizing protease. *Cell* 92:367–380.
16. Sharon M, et al. (2006) 20S proteasomes have the potential to keep substrates in store for continual degradation. *J Biol Chem* 281:9569–9575.
17. Djuranovic S, et al. (2009) Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. *Mol Cell* 34:580–590.
18. Zhang F, et al. (2009) Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol Cell* 34:473–484.
19. Zhang F, et al. (2009) Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol Cell* 34:485–496.
20. Förster F, et al. (2009) An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem Biophys Res Commun* 388:228–233.
21. Tomko RJ, Jr, Funakoshi M, Schneider K, Wang J, Hochstrasser M (2010) Heterohexameric ring arrangement of the eukaryotic proteasomal ATPases: Implications for proteasome structure and assembly. *Mol Cell* 38:393–403.
22. Bohn S, et al. (2010) Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution. *Proc Natl Acad Sci USA* 107:20992–20997.
23. Tian G, et al. (2011) An asymmetric interface between the regulatory and core particles of the proteasome. *Nat Struct Mol Biol* 18:1259–1267.
24. Sanches M, Alves BS, Zanchin NI, Guimaraes BG (2007) The crystal structure of the human Mov34 MPN domain reveals a metal-free dimer. *J Mol Biol* 370:846–855.
25. Riedinger C, et al. (2010) The structure of RPN10 and its interactions with polyubiquitin chains and the proteasome subunit RPN12. *J Biol Chem* 285:33992–34003.
26. Pathare GR, et al. (2011) The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc Natl Acad Sci USA* 109:149–154.
27. Lasker K, et al. (2010) Integrative structure modeling of macromolecular assemblies from proteomics data. *Mol Cell Proteomics* 9:1689–1702.
28. Russel D, et al. (2012) Putting the pieces together: integrative structure determination of macromolecular assemblies. *PLoS Biol*, in press.
29. Alber F, Forster F, Korkin D, Topf M, Sali A (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* 77:443–477.
30. Leitner A, et al. (2010) Probing native protein structures by chemical cross-linking, mass spectrometry and bioinformatics. *Mol Cell Proteomics* 9:1634–1649.
31. Alber F, et al. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450:683–694.
32. Alber F, et al. (2007) The molecular architecture of the nuclear pore complex. *Nature* 450:695–701.
33. Sakata E, et al. (2011) Localization of the proteasomal ubiquitin receptors Rpn10 and Rpn13 by electron cryomicroscopy. *Proc Natl Acad Sci USA*, in press.
34. Lasker K, Topf M, Sali A, Wolfson HJ (2009) Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J Mol Biol* 388:180–194.
35. Förster F, Han BG, Beck M (2010) Visual Proteomics. *Methods Enzymol* 483:215–243.
36. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683.
37. Xie Y, Varshavsky A (2001) RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: A negative feedback circuit. *Proc Natl Acad Sci USA* 98:3056–3061.
38. Seeger M, et al. (2003) Interaction of the anaphase-promoting complex/cyclosome and proteasome protein complexes with multiubiquitin chain-binding proteins. *J Biol Chem* 278:16791–16796.
39. Krogan NJ, et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440:637–643.
40. Kajava AV (2002) What curves alpha-solenoids? Evidence for an alpha-helical toroid structure of Rpn1 and Rpn2 proteins of the 26 S proteasome. *J Biol Chem* 277:49791–49798.
41. Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV (2006) Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol* 4:e267.
42. Takeuchi J, Fujimuro M, Yokosawa H, Tanaka K, Toh-e A (1999) Rpn9 is required for efficient assembly of the yeast 26S proteasome. *Mol Cell Biol* 19:6575–6584.
43. Fukunaga K, Kudo T, Toh-e A, Tanaka K, Saeki Y (2010) Dissection of the assembly pathway of the proteasome lid in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* 396:1048–1053.
44. Wollenberg K, Swaffield JC (2001) Evolution of proteasomal ATPases. *Mol Biol Evol* 18:962–974.
45. Smith DM, et al. (2007) Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol Cell* 27:731–744.
46. Förster F, Lasker K, Nickell S, Sali A, Baumeister W (2010) Towards an integrated structural model of the 26S proteasome. *Mol Cell Proteomics* 9:1666–1677.
47. Lupas A, Baumeister W, Hofmann K (1997) A repetitive sequence in subunits of the 26S proteasome and 20S cyclosome (anaphase-promoting complex). *Trends Biochem Sci* 22:195–196.
48. Effantin G, Rosenzweig R, Glickman MH, Steven AC (2009) Electron microscopic evidence in support of alpha-solenoid models of proteasomal subunits Rpn1 and Rpn2. *J Mol Biol* 386:1204–1211.
49. Nickell S, et al. (2009) Insights into the molecular architecture of the 26S proteasome. *Proc Natl Acad Sci USA* 106:11943–11947.
50. Elsasser S, et al. (2002) Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat Cell Biol* 4:725–730.
51. Gomez TA, Kolawa N, Gee M, Sweredoski MJ, Deshaies RJ (2011) Identification of a functional docking site in the Rpn1 LRR domain for the UBA-UBL domain protein Ddi1. *BMC Biol* 9:33.
52. Tran HJ, Allen MD, Lowe J, Bycroft M (2003) Structure of the Jab1/MPN domain and its implications for proteasome function. *Biochemistry* 42:11460–11465.
53. Ambroggio XI, Rees DC, Deshaies RJ (2004) JAMM: A metalloprotease-like zinc site in the proteasome and signalosome. *PLoS Biol* 2:E2.
54. Isono E, Saito N, Kamata N, Saeki Y, Toh EA (2005) Functional analysis of Rpn6p, a lid component of the 26 S proteasome, using temperature-sensitive rpn6 mutants of the yeast *Saccharomyces cerevisiae*. *J Biol Chem* 280:6537–6547.
55. Yao T, et al. (2006) Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nat Cell Biol* 8:994–1002.
56. Hamazaki J, et al. (2006) A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J* 25:4524–4536.
57. Chen X, Lee BH, Finley D, Walters KJ (2010) Structure of proteasome ubiquitin receptor hRpn13 and its activation by the scaffolding protein hRpn2. *Mol Cell* 38:404–415.
58. Wang Q, Young P, Walters KJ (2005) Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J Mol Biol* 348:727–739.
59. Xie Y, Varshavsky A (2000) Physical association of ubiquitin ligases and the 26S proteasome. *Proc Natl Acad Sci USA* 97:2497–2502.
60. Glickman MH, et al. (1998) A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 94:615–623.
61. Smith DM, Fraga H, Reis C, Kafri G, Goldberg AL (2011) ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. *Cell* 144:526–538.
62. Glynn SE, Martin A, Nager AR, Baker TA, Sauer RT (2009) Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* 139:744–756.
63. Pick E, Hofmann K, Glickman MH (2009) PCI complexes: Beyond the proteasome, CSN, and eIF3 Troika. *Mol Cell* 35:260–264.
64. Gajiwala KS, Burley SK (2000) Winged helix proteins. *Curr Opin Struct Biol* 10:110–116.
65. Collins GA, Tansey WP (2006) The proteasome: A utility tool for transcription? *Curr Opin Genet Dev* 16:197–202.
66. Kouba T, Rutkai E, Karaskova M, Valasek LS (2011) The eIF3c/NIP1 PCI domain interacts with RNA and RACK1/ASC1 and promotes assembly of translation preinitiation complexes. *Nucleic Acids Res*, 10.1093/nar/gkr1083.
67. Zhang N, et al. (2009) Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol Cell* 35:280–290.
68. Deveraux Q, Ustrell V, Pickart C, Rechsteiner M (1994) A 26 S protease subunit that binds ubiquitin conjugates. *J Biol Chem* 269:7059–7061.
69. Piotrowski J, et al. (1997) Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J Biol Chem* 272:23712–23721.
70. Peth A, Uchiki T, Goldberg AL (2010) ATP-dependent steps in the binding of ubiquitin conjugates to the 26S proteasome that commit to degradation. *Mol Cell* 40:671–681.
71. Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* 69:961–1004.
72. Reuter CJ, Kaczowka SJ, Maupin-Furlow JA (2004) Differential regulation of the PanA and PanB proteasome-activating nucleotidase and 20S proteasomal proteins of the haloarchaeon *Haloferax volcanii*. *J Bacteriol* 186:7763–7772.
73. Zhang X, et al. (2004) The N-terminal coiled coil of the *Rhodococcus erythropolis* ARC AAA ATPase is neither necessary for oligomerization nor nucleotide hydrolysis. *J Struct Biol* 146(1–2):155–165.
74. Inobe T, Fishbain S, Prakash S, Matouschek A (2011) Defining the geometry of the two-component proteasome degron. *Nat Chem Biol* 7:161–167.

Supporting Information

Lasker et al. 10.1073/pnas.1120559109

SI Material and Methods.

Sample Preparation, Cryo-EM, and XL/MS. Sample preparation, cryo-EM single particle analysis, and XL/MS were performed as described in ref. 1. In brief, 26S proteasomes were purified from $^{3\times FLAG}$Rpn11 *S. pombe* cells and the vitrified sample was imaged using a Tecnai F20 transmission electron microscope (FEI). From 375,000 single particles, we computed an 8.4-Å resolution map (Fourier-Shell correlation, FSC = 0.5) using XMIPP (2) (Fig. S1). The local resolution at each voxel was computed using the FSC of two reconstructions from each 50% of the particles within a softened sphere (radius = 22 Å) centered at the respective voxel (Fig. 1*B*).

XL/MS. Purified 26S proteasomes from *S. pombe* and *S. cerevisiae* were concentrated to 1 mg/mL. 50 μL of sample were cross-linked with 1 mM DSS (H_{12}/D_{12}, from Creative Molecules). The cross-linking reaction was quenched with 50 mM ammonium bicarbonate, and proteins were reduced, alkylated and digested with 1 μg trypsin (Promega). Cross-linked peptides were enriched using peptide size exclusion chromatography and analyzed by LC-MS/MS using a Thermo LTQ Orbitrap XL mass spectrometer.

MS data were analyzed using xQuest (3). The xQuest search was performed against target and decoy databases of previously identified proteasome proteins with a precursor mass tolerance of 10 ppm. For matching of fragment ions, tolerances of 0.2 Da for common ions and 0.3 Da for cross-link ions were used. FDRs were estimated based on the decoy counts using the software xProphet. All cross-link identifications corresponding to an FDR <5% were accepted.

Comparative Modeling of the Rpn Subunits. For each Rpn subunit, HHpred (4) was used to search for homologous structures (templates), followed by building multiple comparative models with MODELLER-9v8 (5), based on combinations of templates selected manually (Table S3); unaligned termini were truncated. The models were scored by TSVMod (6), a program for predicting the C^{α} root-mean-square error. Models with predicted C^{α}-rmsd errors lower than 8 Å were used for subsequent analyses.

Multiple Fitting of the PCI Subunits. MultiFit (7) was used to simultaneously fit the PCI subunit comparative models into their uniquely localized regions. We ran MultiFit 54 times, each time with a different combination of comparative models, because we could not determine a priori which comparative model to use for each of the PCI subunits. Because the relative orientation between the PCI domain and the α-solenoid domain of each PCI subunit varied widely among its comparative models, we allowed the two domains to move relative to each other as rigid bodies, restrained by a maximum allowed distance. We then clustered the two top-scoring models from all MultiFit runs (108 models in total) according to their PCI-domain centroids. The models clustered into 11 groups. The PCI-domain configuration shared by the three largest clusters (with 15,15, and 12 members, respectively) clearly indicated that the PCI subunits interact via their PCI domains (Fig. S3).

Flexible Fitting of the PCI Subunits, the AAA-ATPase Hexamer, and the CP. We used MDFF (8) to fit available atomic models into the cryo-EM map. All MDFF runs were done with default parameters, using implicit solvent and temperature annealing. The comparative models of the AAA-ATPase hexamer and the CP were first rigidly fitted into the density and then flexibly fitted. To fit the PCI-domain-containing subunits, the PCI domains of the comparative models from the three main clusters found by MultiFit were refined first, while the α-solenoid domains were internally restrained to move as rigid bodies. Additional restraints between PCI-domain β-sheets enforced contacts observed between the PCI domains of Rpn6 monomers in the crystal (PDB ID code 3TXM). In the second step, the α-solenoid restraints were released and the entire subunits were allowed to fit into the cryo-EM map. To prevent the Rpn12 α-solenoid helices from moving into neighboring areas for which no atomic models were available, distance restraints between α-helices were imposed. The accuracy of the flexibly refined models strongly depends on the local resolution of the cryo-EM map (Fig. S1) and the quality of the comparative models (Table S3).

Visualization. All figures were generated using Chimera (9) and VMD (10).

1. Bohn S, et al. (2010) Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution. *Proc Natl Acad Sci USA* 107:20992–20997.
2. Scheres SH, Nunez-Ramirez R, Sorzano CO, Carazo JM, Marabini R (2008) Image processing for electron microscopy single-particle analysis using XMIPP. *Nat Protoc* 3:977–990.
3. Rinner O, et al. (2008) Identification of cross-linked peptides from large sequence databases. *Nat Methods* 5:315–318.
4. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–248.
5. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
6. Eramian D, Eswar N, Shen MY, Sali A (2008) How well can the accuracy of comparative protein structure models be predicted? *Protein Sci* 17:1881–1893.
7. Lasker K, Topf M, Sali A, Wolfson HJ (2009) Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J Mol Biol* 388:180–194.
8. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683.
9. Goddard TD, Huang CC, Ferrin TE (2007) Visualizing density maps with UCSF Chimera. *J Struct Biol* 157:281–287.
10. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38, 27–38.
11. Kajava AV (2002) What curves alpha-solenoids? Evidence for an alpha-helical toroid structure of Rpn1 and Rpn2 proteins of the 26 S proteasome. *J Biol Chem* 277:49791–49798.
12. Förster F, Han BG, Beck M (2010) Visual proteomics. *Methods Enzymol* 483:215–243.
13. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. *Proteins* 64:559–574.
14. Hartmann-Petersen R, Tanaka K, Hendil KB (2001) Quaternary structure of the ATPase complex of human 26S proteasomes determined by chemical cross-linking. *Arch Biochem Biophys* 386:89–94.
15. Richmond C, Gorbea C, Rechsteiner M (1997) Specific interactions between ATPase subunits of the 26 S protease. *J Biol Chem* 272:13403–13411.
16. Takeuchi J, Tamura T (2004) Recombinant ATPases of the yeast 26S proteasome activate protein degradation by the 20S proteasome. *FEBS Lett* 565:39–42.
17. Gorbea C, Taillandier D, Rechsteiner M (2000) Mapping subunit contacts in the regulatory complex of the 26 S proteasome. S2 and S5b form a tetramer with ATPase subunits S4 and S7. *J Biol Chem* 275:875–882.
18. Park S, et al. (2009) Hexameric assembly of the proteasomal ATPases is templated through their C termini. *Nature* 459:866–870.
19. Xie Y, Varshavsky A (2001) RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: A negative feedback circuit. *Proc Natl Acad Sci USA* 98:3056–3061.
20. Funakoshi M, Tomko RJ, Jr, Kobayashi H, Hochstrasser M (2009) Multiple assembly chaperones govern biogenesis of the proteasome regulatory particle base. *Cell* 137:887–899.
21. Wilkinson CR, Wallace M, Seeger M, Dubiel W, Gordon C (1997) Mts4, a non-ATPase subunit of the 26 S protease in fission yeast is essential for mitosis and interacts directly with the ATPase subunit Mts2. *J Biol Chem* 272:25768–25777.

22. Chen C, et al. (2008) Subunit-subunit interactions in the human 26S proteasome. *Proteomics* 8:508–520.
23. Davy A, et al. (2001) A protein-protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep* 2:821–828.
24. Zhang Z, et al. (2000) Structural and functional characterization of interaction between hepatitis B virus X protein and the proteasome complex. *J Biol Chem* 275:15157–15165.
25. Krogan NJ, et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440:637–643.
26. Cagney G, Uetz P, Fields S (2001) Two-hybrid analysis of the *Saccharomyces cerevisiae* 26S proteasome. *Physiol Genomics* 7:27–34.
27. Uetz P, et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–627.
28. Fu H, Reis N, Lee Y, Glickman MH, Vierstra RD (2001) Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome. *EMBO J* 20:7096–7107.
29. Ito T, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* 98:4569–4574.
30. Saeki Y, Toh EA, Kudo T, Kawamura H, Tanaka K (2009) Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle. *Cell* 137:900–913.
31. Gavin AC, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440:631–636.
32. Russell SJ, Sathyanarayana UG, Johnston SA (1996) Isolation and characterization of SUG2. A novel ATPase family component of the yeast 26 S proteasome. *J Biol Chem* 271:32810–32817.
33. Satoh K, Sasajima H, Nyoumura KI, Yokosawa H, Sawada H (2001) Assembly of the 26S proteasome is regulated by phosphorylation of the p45/Rpt6 ATPase subunit. *Biochemistry* 40:314–319.
34. Xie Y, Varshavsky A (2000) Physical association of ubiquitin ligases and the 26S proteasome. *Proc Natl Acad Sci USA* 97:2497–2502.
35. Seeger M, et al. (2003) Interaction of the anaphase-promoting complex/cyclosome and proteasome protein complexes with multiubiquitin chain-binding proteins. *J Biol Chem* 278:16791–16796.
36. Stone M, et al. (2004) Uch2/Uch37 is the major deubiquitinating enzyme associated with the 26S proteasome in fission yeast. *J Mol Biol* 344:697–706.
37. Giot L, et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302:1727–1736.
38. Yao T, et al. (2006) Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nat Cell Biol* 8:994–1002.
39. Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV (2006) Structural organization of the 19S proteasome lid: Insights from MS of intact complexes. *PLoS Biol* 4:e267.
40. Joshi KK, Chen L, Torres N, Tournier V, Madura K (2011) A proteasome assembly defect in rpn3 mutants is associated with Rpn11 instability and increased sensitivity to stress. *J Mol Biol* 410:383–399.
41. Tone Y, et al. (2000) Nob1p, a new essential protein, associates with the 26S proteasome of growing *Saccharomyces cerevisiae* cells. *Gene* 243:37–45.
42. Isono E, Saeki Y, Yokosawa H, Toh-e A (2004) Rpn7 Is required for the structural integrity of the 26 S proteasome of *Saccharomyces cerevisiae*. *J Biol Chem* 279:27168–27176.
43. Gandhi TK, et al. (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* 38:285–293.
44. Pathare GR, et al. (2011) The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc Natl Acad Sci USA* in press
45. Li T, et al. (2001) Identification of two proteins, S14 and UIP1, that interact with UCH37. *FEBS Lett* 488:201–205.
46. Frank J (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State* (Oxford Univ Press, New York) 2nd Ed.
47. Alber F, Kim MF, Sali A (2005) Structural characterization of assemblies from overall shape and subcomplex compositions. *Structure* 13:435–445.
48. Connolly ML (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713.

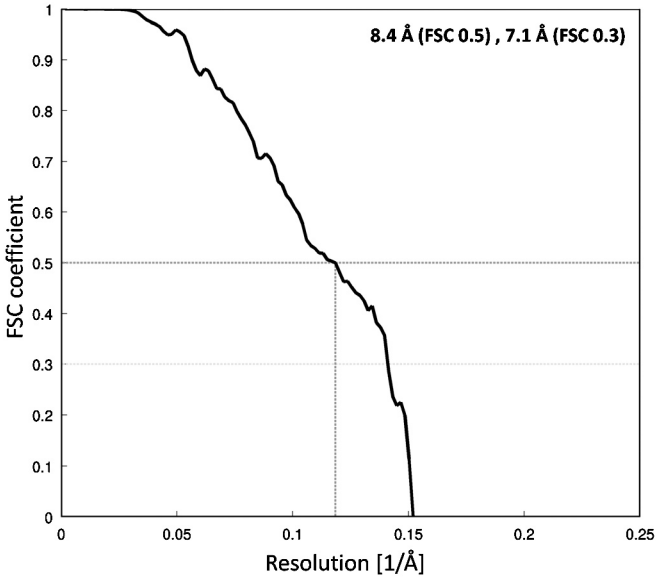

Fig. S1. Resolution assessment of the *S. pombe* 26S proteasome cryo-EM map. The Fourier-Shell Correlation (FSC) coefficient decreases below 0.5 at a resolution of approximately 8.4^{-1} $Å^{-1}$.

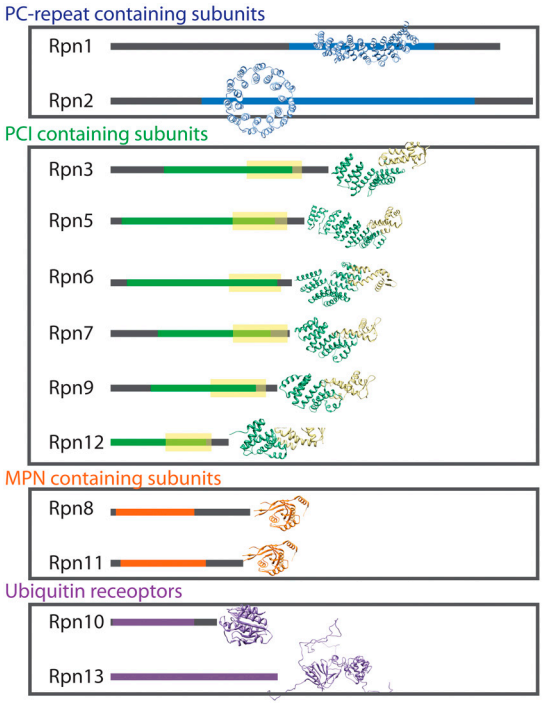

Fig. S2. Comparative modeling of the Rpn subunits. The sequence segments covered by atomic models are indicated. A representative model for each subunit is shown. The PCI domain is indicated in yellow.

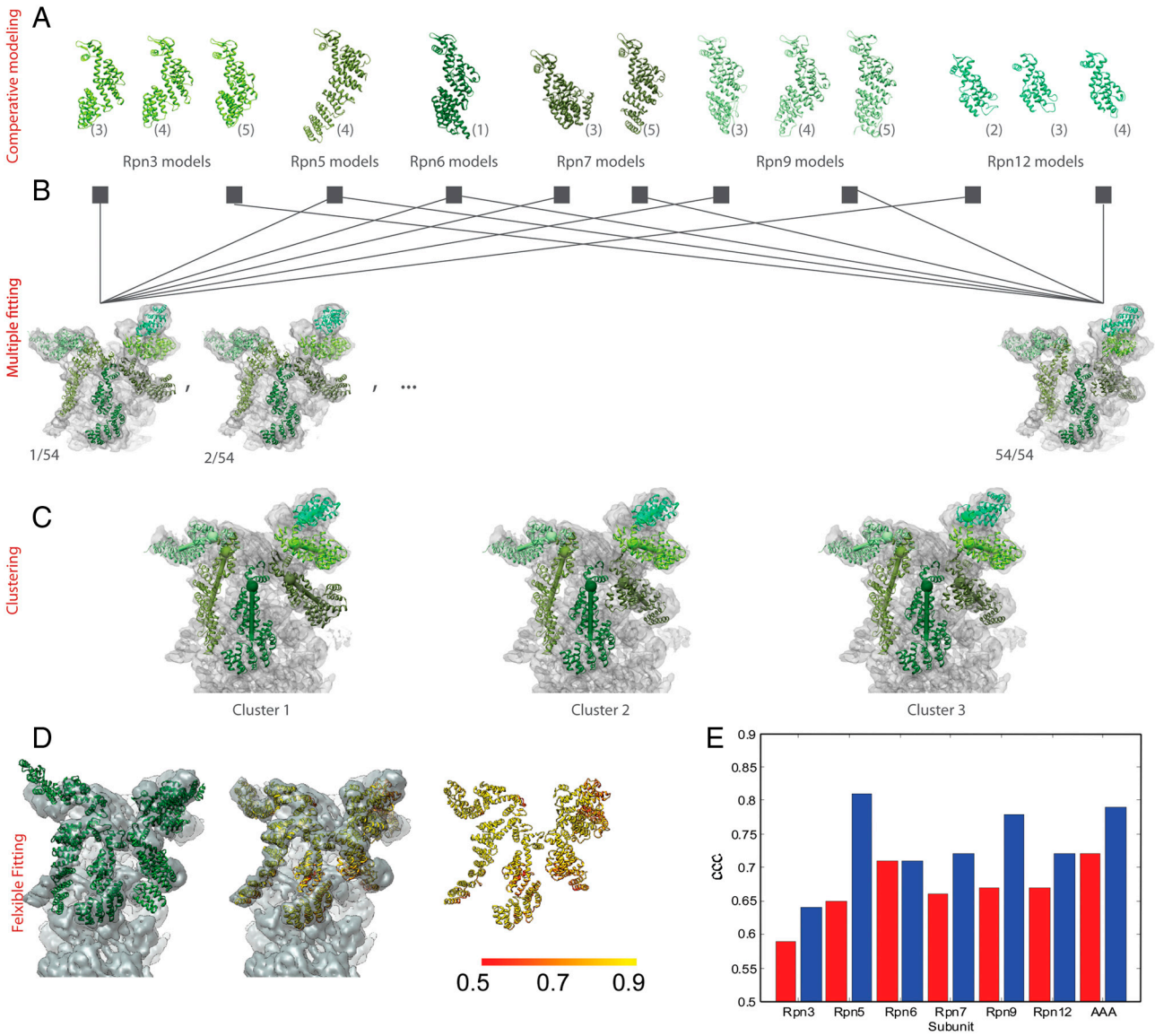

Fig. S3. Multiple fitting of the PCI-domain-containing subunits and flexible fitting. (*A*) Comparative models of the PCI-domain-containing subunits selected for fitting are indexed as in Table S3. (*B*) Fifty-four MultiFit runs, one for each comparative model combination. (*C*) Representatives of the three largest clusters among 108 MultiFit models (two top ranking models for each of the 54 runs). The models were clustered according to their PCI-domain centroids. The PCI-domain centroids, the main principal direction of the α-solenoid domains, as well as the atomic model are shown for each of the three cluster representatives. (*D*) Best-scoring combination of comparative models (green) fitted as rigid bodies into the cryo-EM map that is shown in gray (*Left*); models after flexible fitting by MDFF (8), colored according to the local cross-correlation coefficient of the density of the model and the cryo-EM map (computed analogously as for Fig. 1*B*) shown with (*Middle*) and without the cryo-EM map (*Right*). The local cross-correlation coefficient (colorcode explained in legend) is an indicator for the local confidence of the atomic models. The local correlation suggests that the accuracy of models is higher for the PCI domains than for the solenoid domains, in particular those of Rpn3 and Rpn7 (second and third from the right, respectively). (*E*) Cross-correlation coefficients (CCC, computed by UCSF Chimera) (9) of PCI-domain-containing subunits and the AAA-ATPase hexamer (Fig. 5) before (red) and after flexible fitting (blue). For all subunits, with the exception of Rpn6, whose initial model based on the *D. melanogaster* Rpn6 crystal structure (PDB ID code 3TXM), the correlation coefficients improve notably.

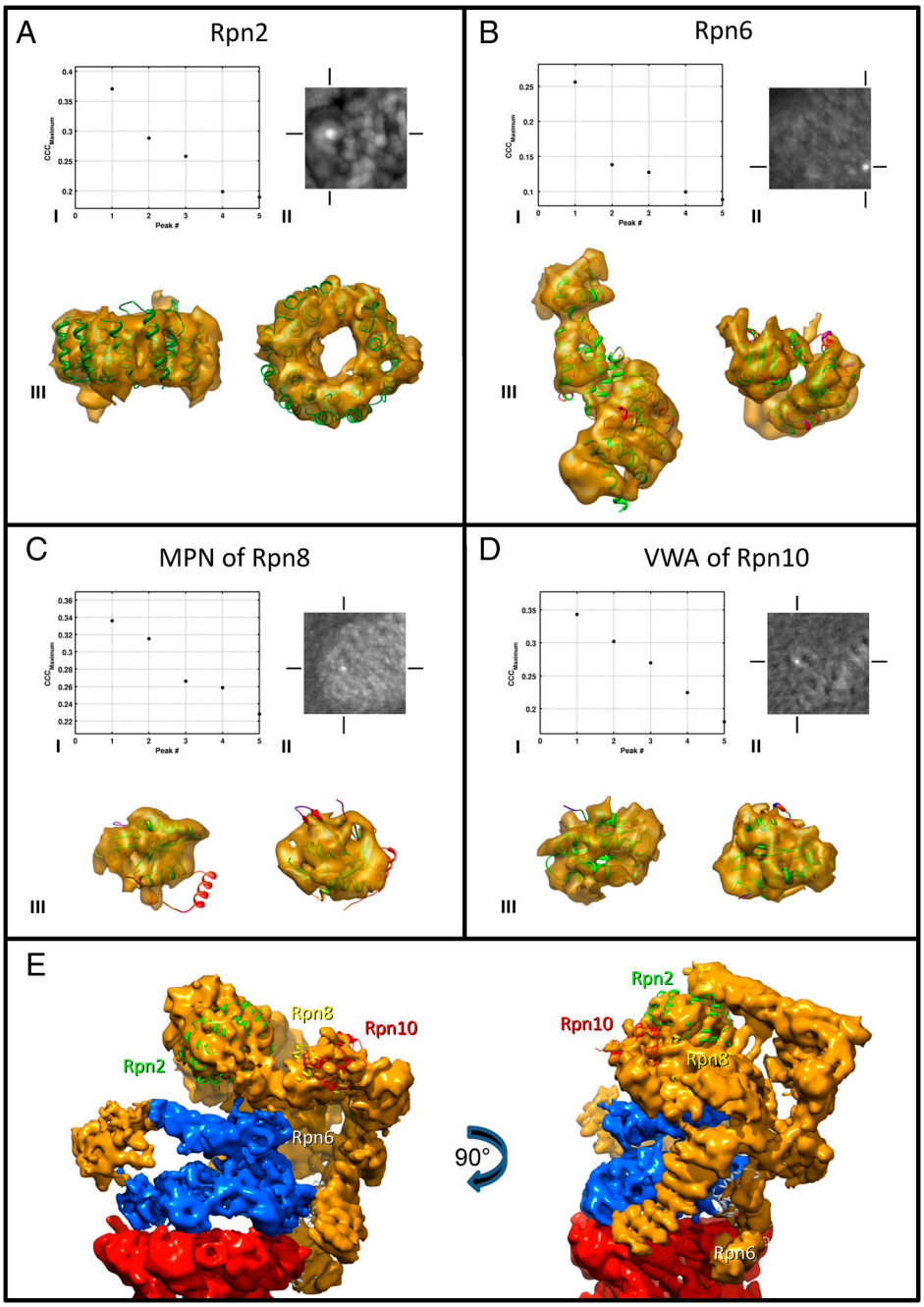

Fig. S4. Fitting of Rpn2, Rpn6, Rpn8, and Rpn10 into the cryo-EM map. (*A*) An atomic model of the Rpn2 LRR domain (residues 309–712) (11) was fitted into the B-factor corrected density ($B = -1{,}000$ Å^2) using MOLMATCH (12). The first five local maxima of the local correlation function are plotted in descending order (I) and a slice of the correlation function is shown with its maximum indicated (II). Moreover, the atomic model (green) is superposed onto the density (gold) identified by the correlation maximum (III). (*B*) Same for the Rpn6 crystal structure (PDB ID code 3TXM), which is colored according to its B factor, ranging from green to red in III. (*C* and *D*) Same for crystal structure of the Rpn8 MPN domain (PDB ID code 2O95), and the von Willenbrand factor A (VWA), domain of Rpn10 (PDB ID code 2X5N). (*E*) The fitted subunits placed in the cryo-EM density, seen from two different views.

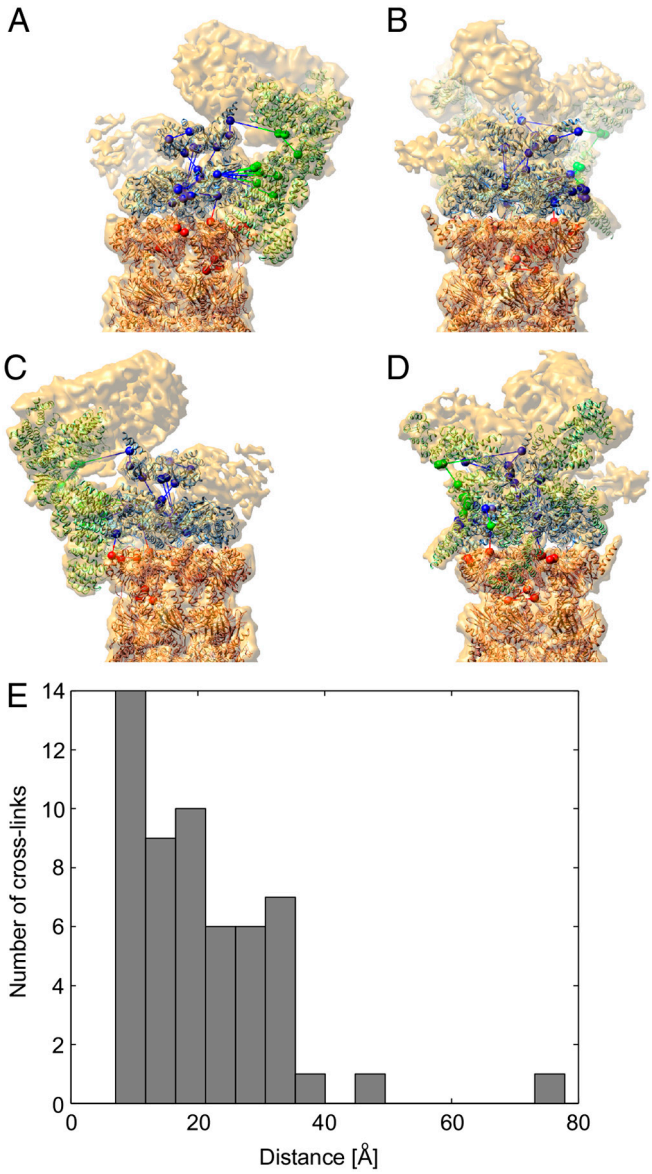

Fig. S5. Cross-links in the atomic model of the 26S proteasome. (*A–D*) The cross-links (Table S1) are plotted in the resulting models, seen from four different views. The C^{α} atoms of the cross-linked residues pairs are displayed as spheres that are connected by cylinders. The coloring is chosen according to subunit type (nonATPases: green, ATPases: blue, CP: red). Accordingly, cylinders connecting residues from different subunit types have two colors. Of 55 cross-linked C^{α} pairs, only two are more than 40 Å apart from each other, involving Rpt4 and Rpn5 (48 Å), as well as Rpn9 and Rpt3 (77 Å). This would suggest a false-positive rate <4% in the cross-linking data, which is predicted to be better than 5%. Moreover, the Rpt4-Rpn5 cross-link could also be due to motion of the Rpt4/Rpt5 coiled-coil. (*E*) Histogram of C^{α}-C^{α} distances for the cross-links.

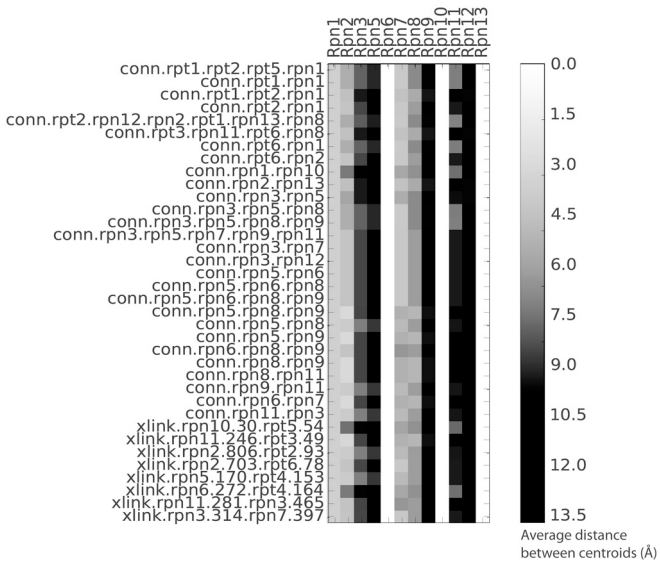

Fig. S6. Ensemble analysis. To assess the redundancy of the restraints, we omitted one proteomics or residue-specific cross-linking restraint at a time and recalculated $5 \cdot 10^5$ top-scoring solutions violating at most five restraints for each restraint subset. We then compared the resulting "jack-knife" ensembles with the union of the refined representative models from each one of the three clusters computed from all restraints (union model). The rows of the matrix correspond to the individual omitted restraints and the columns correspond to the individual Rpn subunits. Each bin indicates the sensitivity of the subunit localization to the presence of a specific restraint, quantified by the average centroid distance between the union and jack-knife subunit models. Even after omitting one restraint at a time, the jack-knife ensembles remain similar to the union model, with the maximum sensitivity of approximately 13.5 Å.

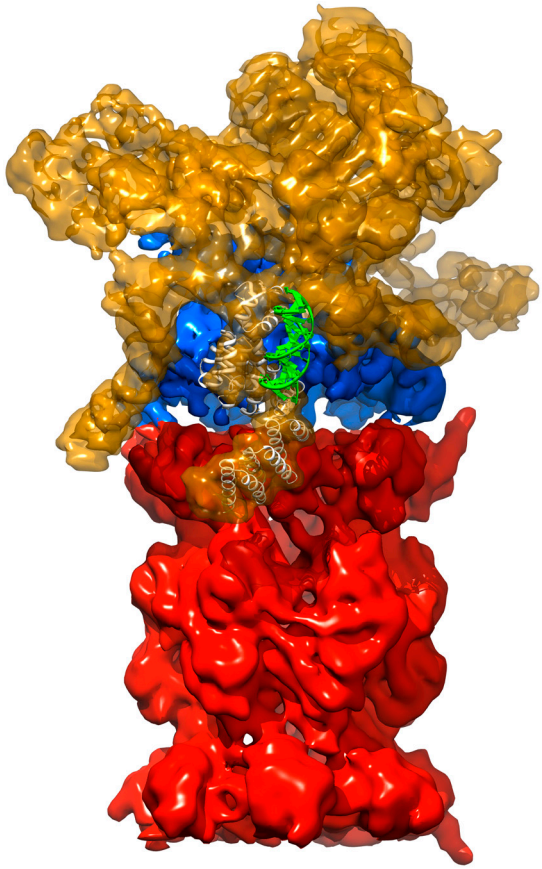

Fig. S7. Putative DNA binding by Rpn6. The transcription factor E2F-DP in complex with DNA (PDB ID code 1CF7) was structurally aligned to the winged helix motif from Rpn6 (white) using MUSTANG (13). The bound DNA molecule (green) does not yield major clashes with surrounding subunits.

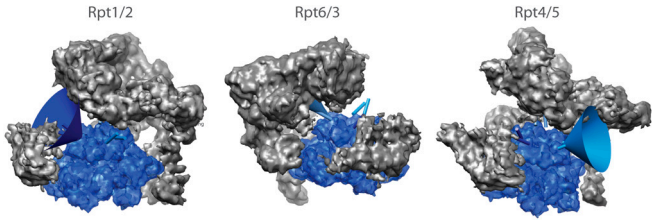

Fig. S8. Analysis of the space accessible for the AAA-ATPase coiled coils pairs. The maximum opening angles of cones originating from the coiled-coil hinge residues corresponding to Pro62 in PAN were computed so that the coiled coils do not penetrate neighboring densities (Rpns: gray, remaining AAA-domains: blue). The opening angles were approximately 40° for the Rpt1/2 coiled-coil, 15° for Rpt6/3, and 40° for Rpt4/5 (blue cylinders).

Table S1. Cross-linked lysine residues in the 26S proteasomes from *S. pombe* and *S. cerevisiae*

Residue 1	Residue 2	Species
IntraCP (6 from *S. pombe* and 1 from *S. cerevisiae*)		
α1:102	α2:91	*S. pombe*
α1:102	α2:98	*S. pombe*
α1:57	α7:177	*S. pombe*
α2:176	α3:54	*S. pombe*
α2:176	α3:50	*S. pombe*
α6:115	α7:89	*S. pombe*
α1:102	α2:91	*S. cerevisiae*
CP-ATPase (4 from *S. pombe* and 2 from *S. cerevisiae*)		
α1:169	Rpt4:297	*S. pombe*
α2:176	Rpt6:390	*S. pombe*
α4:168	Rpt1:422	*S. pombe*
α4:168	Rpt1:426	*S. pombe*
α4:176	Rpt1:415	*S. cerevisiae*
α5:66	Rpt1:426	*S. cerevisiae*
Intra ATPase (32 from *S. pombe* and 8 from *S. cerevisiae*)		
Rpt1:422	Rpt2:357	*S. pombe*
Rpt1:426	Rpt2:357	*S. pombe*
Rpt1:68	Rpt2:93	*S. pombe*
Rpt1:124	Rpt3:207	*S. pombe*
Rpt1:112	Rpt4:38	*S. pombe*
Rpt1:120	Rpt4:205	*S. pombe*
Rpt1:102	Rpt5:172	*S. pombe*
Rpt1:124	Rpt5:124	*S. pombe*
Rpt1:124	Rpt5:172	*S. pombe*
Rpt1:207	Rpt5:413	*S. pombe*
Rpt1:344	Rpt5:432	*S. pombe*
Rpt2:264	Rpt6:84	*S. pombe*
Rpt2:264	Rpt6:90	*S. pombe*
Rpt2:45	Rpt6:90	*S. pombe*
Rpt3:378	Rpt4:297	*S. pombe*
Rpt3:207	Rpt5:124	*S. pombe*
Rpt3:299	Rpt6:357	*S. pombe*
Rpt3:299	Rpt6:390	*S. pombe*
Rpt4:10	Rpt5:51	*S. pombe*
Rpt4:10	Rpt5:54	*S. pombe*
Rpt4:71	Rpt5:124	*S. pombe*
Rpt4:47	Rpt5:77	*S. pombe*
Rpt4:196	Rpt5:217	*S. pombe*
Rpt4:196	Rpt5:350	*S. pombe*
Rpt4:205	Rpt5:124	*S. pombe*
Rpt4:24	Rpt5:64	*S. pombe*
Rpt4:297	Rpt5:217	*S. pombe*
Rpt4:318	Rpt5:217	*S. pombe*
Rpt4:38	Rpt5:64	*S. pombe*
Rpt4:38	Rpt5:72	*S. pombe*
Rpt4:38	Rpt5:77	*S. pombe*
Rpt4:41	Rpt5:64	*S. pombe*
Rpt1:124	Rpt5:124	*S. cerevisiae*
Rpt3:186	Rpt4:164	*S. cerevisiae*
Rpt3:299	Rpt6:390	*S. cerevisiae*
Rpt4:14	Rpt5:44	*S. cerevisiae*
Rpt4:318	Rpt5:212	*S. cerevisiae*
Rpt4:37	Rpt5:75	*S. cerevisiae*
Rpt4:41	Rpt5:75	*S. cerevisiae*
Rpt4:71	Rpt5:124	*S. cerevisiae*

Residue 1	Residue 2	Species
nonATPase-ATPase (12 *S. pombe*, 6 *S. cerevisiae*)		
Rpn10:30	Rpt5:54	*S. pombe*
Rpn11:246	Rpt3:49	*S. pombe*
Rpn11:276	Rpt3:49	*S. pombe*
Rpn11:276	Rpt3:94	*S. pombe*
Rpn11:281	Rpt3:49	*S. pombe*
Rpn2:806	Rpt2:93	*S. pombe*
Rpn2:703	Rpt6:78	*S. pombe*
Rpn2:703	Rpt6:80	*S. pombe*
Rpn5:170	Rpt4:153	*S. pombe*
Rpn5:200	Rpt4:153	*S. pombe*
Rpn5:204	Rpt4:153	*S. pombe*
Rpn6:272	Rpt4:164	*S. pombe*
Rpn11:153	Rpt6:56	*S. cerevisiae*
Rpn2:703	Rpt2:104	*S. cerevisiae*
Rpn2:703	Rpt6:79	*S. cerevisiae*
Rpn5:203	Rpt4:153	*S. cerevisiae*
Rpn6:185	Rpt3:193	*S. cerevisiae*
Rpn9:74	Rpt5:44	*S. cerevisiae*
intra nonATPase (3 *S. pombe*, 5 *S. cerevisiae*)		
Rpn11:281	Rpn3:465	*S. pombe*
Rpn3:314	Rpn7:397	*S. pombe*
Rpn3:318	Rpn7:69	*S. pombe*
Rpn2:625	Rpn3:463	*S. cerevisiae*
Rpn2:925	Rpn5:7	*S. cerevisiae*
Rpn2:178	Rpn9:111	*S. cerevisiae*
Rpn3:277	Rpn7:397	*S. cerevisiae*
Rpn5:286	Rpn9:135	*S. cerevisiae*

For each pair of cross-linked lysines, the subunit name and the residue indices in the respective subunit from *S. pombe* are listed. ATPase refers to the AAA-ATPase hexamer.

Table S2. Protein interactions in the RP involving six or less subunits

Subcomplex	Technique	References	Species
Rpt1, Rpt2	IVB, x-link	(14–17)	sc,h
Rpt1, Rpt3	x-link	(14)	h
Rpt1, Rpt2, Rpt3, Rpt6	IVB	(15)	h
Rpt1, Rpt2, Rpt5, Rpn1	IVPD	(18)	sc
Rpt1, Rpn1	IVB	(17, 19)	h
Rpt1, Rpt2, Rpn1	IVB	(17, 20)	h
Rpt2, Rpt6	x-link	(14)	h
Rpt2, Rpn1	IVPD, 2h, IVB	(17, 21–23)	h
Rpt2, alpha4	IVPD, 2h	(22, 24)	h
Rpt2, Rpn12, Rpn2, Rpt1, Rpn13, Rpn8	IVPD	(25)	sc
Rpt3, Rpt4	2h	(22, 26, 27)	sc
Rpt3, Rpt5	2h, x-link	(14, 26–29)	sc,h
Rpt3, Rpt6	2h, IVB, IVPD	(15, 22, 26, 28, 30)	sc,h
Rpt3, Rpn11, Rpt6, Rpn8	IVPD	(31)	sc
Rpt4, Rpt5	2h, IVB, x-link, IVPD	(14, 15, 20, 23, 28, 29)	w,sc,h
Rpt4, Rpt6	Y2h, x-link	(14, 22, 32)	sc,h
Rpt4, alpha4	Y2h	(22, 23)	sc,w
Rpt4, alpha6	x-link	(14)	h
Rpt6, alpha2	x-link	(33)	p
Rpt6, Rpn1	IVB	(34)	sc
Rpt6, Rpn2	x-link	(14)	h
Rpn1, Rpn10	IVB	(19, 35)	sc
Rpn1, Ubp6	IVB	(36)	sp
Rpn2, Rpn13	2h, IVPD	(29, 37, 38)	dm,sc,sp
Rpn3, Rpn5	x-link	(39)	sc
Rpn3, Rpn5, Rpn8	x-link	(39)	sc
Rpn3, Rpn5, Rpn8, Rpn9	x-link	(39)	sc
Rpn3, Rpn5, Rpn7, Rpn9, Rpn11	x-link	(39)	sc
Rpn3, Rpn7	2h	(23, 28, 29)	sc,w,h
Rpn3, Rpn11	IVB	(40)	sc
Rpn3, Rpn12	2h	(22, 28, 41)	sc
Rpn3, Rpn15	x-link	(39)	sc
Rpn5, Rpn6	2h	(22, 29, 42)	sc
Rpn5, Rpn6, Rpn8	WCMS	(39)	sc
Rpn5, Rpn6, Rpn8, Rpn9	WCMS	(39)	sc
Rpn5, Rpn8, Rpn9	WCMS	(39)	sc
Rpn5, Rpn8	IVPD	(43)	h
Rpn5, Rpn9	2h	(28, 37)	sc,dm
Rpn6, Rpn7	IVB	(44)	dm
Rpn6, Rpn8, Rpn9	WCMS	(39)	sc
Rpn7, Rpn15	x-link	(39)	sc
Rpn8, Rpn9	2h	(22, 23)	sc
Rpn8, Rpn11	2h	(23, 26, 28, 37)	sc,dm
Rpn9, Rpn11	2h	(23, 28)	sc
Rpn10, UCH37	IVPD	(36)	sp
Rpn12, UCH37	IVB, 2h	(45)	h
Rpn13, UCH37	IVPD, IVB	(38, 43)	sp

Protein–protein interactions were compiled from two-hybrid screens (2h), chemical cross-linking (x-link), in vitro binding assays (IVB), different in vivo pull-down experiments (IVPD), and whole complex mass spectrometry (WCMS). The data were obtained from 26S proteasomes of *Homo sapiens* (h), *Saccharomyces cerevisiae* (sc), *Schizomycetes pombe* (sp), *Sus barbatus* (p), *Drosophila melanogaster* (dm), and *Caenorhabditis elegans* (w).

Table S3. Comparative modeling of the PCI-domain-containing subunits Rpn3/5/6/7/9/12

Model, used for multiple fitting (Y/N)	Templates(PDB code, chain id, region, aligned region on target)	Modeled region (Residue range), Coverage (%)	TSVMod Predicted Cα rmsd (Å) (33)	TSVMod Predicted native overlap at 3.5 Å (%)
Rpn3				
Rpn3_1, N	1hz4, A, 1–373, 50–453	67–448, 77%	9.2	41
Rpn3_2, N	1hz4, A, 1–373, 50–453 3q15, A, 4–376, 51–411 1qqe, A, 2–288, 73–369 2ifu, A, 2–307, 67–402 3u4t, A, 1–268, 73–368	59–448, 78%	11.0	33
Rpn3_3, Y	3txm, A, 87–353, 167–426 1qqe, A, 37–265, 125–347	124–425, 60%	5.3	57
Rpn3_4, Y	3txm, A, 87–353, 167–426 1qqe, A , 37–265, 125–347 3chm,A, 26–158, 292–427	124–427, 60%	6.0	60
Rpn3_5, Y	3txm, A, 87–353, 167–426 3chm, A, 26–158, 292–427 1hz4, A, 55–287, 129–353	128–424, 60%	5.5	60
Rpn5				
Rpn5_1, N	2y4t, A, 41–446, 2–432	1–423, 95%	14.2	11
Rpn5_2, N	2y4t, A, 41–446, 2–432 1hz4, A, 2–373, 1–354 1qqe, A, 1–292, 10–323 3q15, A, 69–375, 1–255	1–423, 95%	14.8	8
Rpn5_3, N	3txm, A, 3–352, 45–399 2ifu, A, 2–231, 23–248	23–399, 85%	11.0	13
Rpn5_4, Y	3txm, A, 3–352, 45–399 1ufm, A, 5–78, 329–402 3chm, A, 4–158, 225–401 2ifu, A, 2–231, 23–248 1qqe, A, 43–228, 58–241	24–398, 85%	7.2	57
Rpn5_5, N	3txm, A, 3–352, 45–399 1ufm, A, 5–78, 329–402 2ifu, A, 2–231, 23–248	24–398, 85%	8.9	34
Rpn7				
Rpn7_1, N	1qqe, A, 2–288, 39–318	39–356, 77%	13.7	30
Rpn7_2, N	1qqe, A, 2–288, 39–318 3gw4, A, 2–202, 17–273 1elr, A, 1–129,1 07–256 3gyz, A, 2–150, 75–239 2xcb, A, 1–142, 86–246 1ufm, A, 6–84, 304–382	22–356, 82%	14.8	8
Rpn7_3, Y	3txm, 90–352, 112–373 3chm, A, 25–161, 230–378 2ifu, A, 53–259,3 9–256	37–356, 78%	6.2	62
Rpn7_4, N	3txm, A, 90–352, 112–373 1ufm, A, 12–79, 310–377 1qqe, A, 57–275, 92–305 2ifu, A, 53–259, 39–256	39–373, 82%	9.7	28
Rpn7_5, Y	3txm, A, 90–352, 112–373 1ufm, A, 12–79, 310–377 3chm, A, 25–161, 230–378 1qqe, A, 57–275, 92–305 2ifu, A, 53–259, 39–256	87–346, 64%	4.1	74
Rpn9				
Rpn9_1, N	2ifu, A, 30–307, 1–278 1wi9, A, 4–72, 274–346	1–324, 85%	9.0	35
Rpn9_2, N	2ifu, A, 30–307, 1–278 3chm, A, 2–169, 170–353 3edt, B, 19–282, 1–263 3as5, A, 2–185, 1–200 2fo7, A, 1–136, 25–192 1wi9, A, 4–72, 274–346 1ufm_A, 4–84, 260–349	1–342, 87%	10.6	34
Rpn9_3, Y	3chm, A, 24–169, 201–353 3txm, A, 25–352, 20–340 3edt, B, 22–199, 4–189	1–339, 89%	6.5	58
Rpn9_4, Y	3txm, A, 25–352, 20–340 1ufm, A, 14–78, 270–343 2ifu, A, 83–261, 85–267	1–339, 89%	4.5	70
Rpn9_5, Y	3chm, A, 24–169, 201–353	1–399, 89%	6.8	57

Model, used for multiple fitting (Y/N)	Templates(PDB code, chain id, region, aligned region on target)	Modeled region (Residue range), Coverage (%)	TSVMod Predicted Cα rmsd (Å) (33)	TSVMod Predicted native overlap at 3.5 Å (%)
	3txm, A, 25–352, 20–340 1ufm, A, 14–78, 270–343 2ifu, A, 83–261, 85–267 3edt, B, 22–199, 4–189			
Rpn12				
Rpn12_1, N	1rz4, A, 1–222, 1–269	18–262, 90%	9.6	37
Rpn12_2, Y	1rz4, A, 1–222, 1–269 3txm, A, 172–351, 52–222 1umf, A, 11–79, 166–227	51–221, 63%	5.4	65
Rpn12_3, Y	1rz4, A, 1–222, 1–269 3txm, 172–351, 52–222	51–211, 59%	3.7	78
Rpn12_4, Y	3txm, 172–351, 52–222 1umf, A, 11–79, 166–227	53–221, 63%	3.6	79

In addition, we used a comparative model of the AAA-ATPase hexamer based on the PAN structures (PDB ID codes 3H43 and 3H4M). Comparative models for Rpn1/2/8/11 were also built but not used here. The comparative model for Rpn6 was calculated in a previous study (44). Coverage indicates the percentage of target residues in the modeled region. The TSVmod program was used to predict model accuracies in terms of the C^{α} rmsd and native overlap at 3.5-Å cutoff (33).

Table S4. Representation of the *S. pombe* 26S proteasome RP subunits

Subunit Name, UniProt code, UniProt name	No. of residues	Coarse-grained representation No. of beads (50 residues per bead), Rigid (R)/Flexible (F), No. of sampled localizations	Atomic representation First–last residue in a segment
Rpt1, O42931, PRS7_SCHPO	438	9, R, 1	77–427
Rpt2, P36612, PRS4_SCHPO	448	9, R, 1	96–441
Rpt3, O74894, PRS6B_SCHPO	389	8, R, 1	36–380
Rpt4, O74445, PRS10_SCHPO	388	8, R, 1	68–433
Rpt5, O14126, PRS6A_SCHPO	438	9, R, 1	38–383
Rpt6, P41836, PRS8_SCHPO	403	8, R, 1	49–394
Rpn1, P87048, RPN1_SCHPO	891	18, F, 39,269	None
Rpn2, O74762, RPN2_SCHPO	965	19, F, 68,261	309–712
Rpn3, O42897, RPN3_SCHPO	497	10, F, 31,456	124–425
Rpn5, Q9UTM3, RPN5_SCHPO	443	9, F, 12,994	24–398
Rpn6, Q9P7S2, RPN6_SCHPO	421	8, R, 1	1–374
Rpn7, Q10335, RPN7_SCHPO	409	8, F, 157	87–346
Rpn8, O74440, RPN8_SCHPO	324	6, F, 5,023	15–186
Rpn9, Q9US13, RPN9_SCHPO	381	8, F, 39,230	1–399
Rpn10, O94444, RPN10_SCHPO	243	5, F, 1	2–190
Rpn11, P41878, RPN11_SCHPO	308	6, F, 2,044	None
Rpn12, P50524, RPN12_SCHPO	270	5, F, 5,955	53–221
Rpn13, Q9USM1, ADRM1_SCHPO	388	8, R, 1	None

In the coarse-grained representation, each RP subunit was represented as a string of beads, each bead corresponding to an approximately 50-residue fragment. If the atomic structure of a subunit was reliably predicted by comparative modeling, the beads were positioned rigidly (R) to mimic the shape of the atomic model, otherwise, the string of beads was allowed to flex (F). In the atomic representation, atomic coordinates were used for segments with available comparative models.

Table S5. Restraint description

Information	Restraint	Description	Restraint parameters	Violation cutoff	Sampling stage
8.4-Å cryo-EM density map of the *S. pombe* 26S proteasome, Fig. 1	Quality-of-fit into a density map	The fit of a model into an assembly density map was quantified by a cross-correlation measure between the assembly density and the model smoothed to the resolution of the map (46).	Gaussian kernel σ of 4.2 Å	N/A	Subunit fitting
Chemical cross-links from XL/MS, Table S1	Intersubunit residue–residue distance restraint	In the coarse-grained representation, each residue–residue proximity is translated into subunit–subunit proximity using the subomplex connectivity restraint. In the atomic representation, each residue-residue proximity is encoded as a pairwise distance restraint between the cross-linked atoms.	A distance restraint is defined as a lower bound harmonic restraint between pairs of beads with the following parameters: f is the sphere distance (surface-to-surface distance) between the two beads, f_0 is set to 20 Å to allow for long linkers, and K is set to 0.0236, corresponding to the standard deviation of 5 Å.	A restraint is violated when its value is exceeds 1.18, corresponding to two standard deviations of the upper bound. The maximal number of allowed violations is five.	Subunit localization and subunit fitting
Proteomics data, Table S2	Subcomplex connectivity	Each direct interaction or proximity between subunits is encoded as a graph of nodes connected by edges. Each node represents a bead of a subunit, and each edge represents a possible interaction between beads. Edges are generated for all possible interactions that might be implied by the corresponding experiment and for pairs of beads representing sequential fragments of the same subunit. An edge imposes a putative distance restraint, scoring well when the two potentially interacting beads are actually interacting in the assessed assembly configuration, and scoring poorly when the beads are too far apart to interact. For each physical interaction graph, the best-scoring set of edges needed to connect all nodes is chosen by calculating a minimal spanning tree (MST) on the composite graph. The value of the restraint is the sum of all edge distance restraints in the MST.	A distance restraint is defined as a lower bound harmonic restraint between pairs of beads (or atoms) with the following parameters: f is the sphere distance between the two beads, f_0 is set to 10 Å for direct interaction and to 10 Å plus maximal linker length for cross-links, and K is set to 0.0236, corresponding to the standard deviation of 5 Å.	A restraint is violated when its value exceeds $(n-1)*m$, where n is the number of subunits restrained and m is 1.18; $n-1$ is the number of edges in the MST. The maximal number of allowed violations is four.	Subunit localization and subunit fitting
Bioinformatics	Radius-of-gyration (ROG) restraint	A harmonic upper bound function is imposed between the predicted ROG of the subunit (47) and the actual ROG. Because a subunit may adapt a nonglobular geometry, the predicted ROG is scaled by 1.6.	f_0 is the predicted ROG multiplied by 1.6, f is the actual ROG, and K is set to 1.	A restraint is violated if its value exceeds 1.18. The maximal number of allowed violations is 0.	Subunit localization

Information	Restraint	Description	Restraint parameters	Violation cutoff	Sampling stage
Bioinformatics	Excluded volume restraint	To ensure a subunit does not penetrate another, an excluded volume restraint is applied to pairs of subunits. The restraint is defined as a sum of upper bound distance restraints between intersubunit pairs of beads.	A distance restraint is defined as an upper bound harmonic restraint between pairs of beads with the following parameters:f is the sphere distance between the two beads, f_0 is set to 0 Å, and K is set to 1.	A restraint is violated when its value exceeds $n_1 * n_2 * r * m$, where $n_1(n_2)$ is the number of beads of the first (second) subunit, r is the fraction of bead pairs allowed to overlap, here set to 0.1, and m 1.18. The maximal number of allowed violations is 13.	Subunit localization
Bioinformatics	Geometric complementarity restraint	A Connolly surface is calculated for each component (48). Interacting surfaces were scored by summing a reward for the number of surface atom pairs within a distance cutoff and a penalty for all clashing pairs of atoms (7).	Default parameters were used (48), except for the shell widths that were twice as wide to allow for uncertainty in comparative models.	A restraint is violated when its value is larger than 1. The maximal number of allowed violations is 5.	Subunit fitting

Many of the restraints, such as the subcomplex connectivity, radius of gyration, and intersubunit distance restraint, are defined as a sum of harmonic functions, whereas other restraints, such as the quality-of-fit and geometric complementarity restraints, are defined as correlation functions. A harmonic function is $\frac{1}{2}K(f - f_0)^2$, where f is the restrained feature and K is the parameter that determines the strength of the restraint. For an upper bound, the score is 0 for $f < f_0$; for a lower bound, the score is 0 for $f > f_0$.