Insights into the molecular architecture of the 26S proteasome

Stephan Nickellb, Florian Becka, Sjors H. W. Scheresb, Andreas Korinekea, Friedrich Förstera,c, Keren Laskerd, Oana Mihalachea, Na Suna, István Nagya, Andrej Sali, Jürgen M. Plitzkoe, Jose-Maria Carazo, Matthias Mann, and Wolfgang Baumeistera,b

aMax Planck Institute of Biochemistry, D-82152 Martinsried, Germany; bCentro Nacional de Biotecnología–Consejo Superior de Investigaciones Científicas, Cantoblanco, 28049, Madrid, Spain; cDepartment of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158-2330; and dBlavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Communicated by Alexander Varshavsky, California Institute of Technology, Pasadena, CA, May 12, 2009 (received for review March 9, 2009)

Cryoelectron microscopy in conjunction with advanced image analysis was used to analyze the structure of the 26S proteasome and to elucidate its variable features. We have been able to outline the boundaries of the ATPase module in the “base” part of the regulatory complex that can vary in its position and orientation relative to the 20S core particle. This variation is consistent with the “wobbling” model that was previously proposed to explain the role of the regulatory complex in opening the gate in the α-rings of the core particle. In addition, a variable mass near the mouth of the ATPase ring has been identified as Rpn10, a multienzyme receptor, by correlating the electron microscopy data with quantitative mass spectrometry.

Results and Discussion

MS Analysis of Purified 26S Proteasomes. In the case of the 26S proteasome, progress in elucidating its structure has been hampered by the complexity of the system, its variability, and its fragility. 26S proteasome preparations from Drosophila melanogaster embryos (11, 12) appear biochemically and structurally homogeneous and have a well-defined complement of subunits. For an assessment of the subunit stoichiometry of the purified complex with the 26S proteasome. In a second step the resulting intensities were standardized relative to the subunits of the 20S core particle (standard deviation is 0.24). All canonical subunits of the 20S proteasome were found to be present in equimolar amounts. In contrast, paralogs of Rpt3 and Rpt6 (indicated by asterisk) were detected in far substoichiometric quantities (<1:100), suggesting that these subunits are exchangeable. A subunit of the 26S base complex, Rpn10, was determined with a relative intensity of 1:4, indicative of a transient interaction with the 26S proteasome.


The authors declare no conflict of interest.

1To whom correspondence should be addressed at: Max Planck Institute of Biochemistry, Department of Structural Biology, Am Klopferspitz 18, D-82152 Martinsried, Germany. E-mail: baumeist@biochem.mpg.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0905081106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0905081106
complexes are abundant (12). However, suspension in a thin aqueous film before cryofixation causes partial disassembly even when great care is taken to avoid depletion of ATP, which is required for stability; this may be the result of multiple interactions with the water–air interface. Chemical fixation using glutaraldehyde or other cross-linkers is effective in reducing disintegration, but visual inspection of averages derived from such preparations indicates that some structural alterations do occur; therefore, we have not used any of these stabilizing measures.

**Determination of the 26S Proteasome Density by Cryo-EM.** The intrinsic heterogeneity of 26S preparations and the fragility of the complexes give rise to a disturbing degree of variation in cryo-EM sample preparations. As a consequence, it is necessary to collect a large number of particle images, to be able to classify the particles into sufficiently populated classes representing the distinct states present in the sample. A recently developed automated image acquisition procedure (14) has enabled us to collect large data sets of 26S proteasomes, despite the relatively low abundance of double-capped holocomplexes.

A first 3D reconstruction (Fig. 2) was obtained by using a standard angular refinement procedure as implemented in the XMIPP software package (15). To minimize any model bias, we used a plain cylinder model as an initial reference to assign optimal projection directions and in-plane transformations to all experimental particles, and this model was iteratively refined until no further improvements were observed. The quality of the resulting reconstruction (left 2 columns in Fig. 2) can be assessed by using the low-pass filtered (2-nm cut-off) crystal structure of the 20S CP as an internal standard as well as for optimizing the isosurface threshold. The included total volume corresponds to a molecular mass of 2.5 MDa, which is in good agreement with the proteomics analysis (11). Moreover, the fit of the 20S CP into the cryo-EM density map, found after an extensive correlation search, is excellent [cross-correlation function (CCC) = 0.69]. Therefore, we conclude that the overall average structure with a nominal resolution of 2.5 nm is accurate. Nevertheless, given the dynamics of this molecular machine, we expect that significant local variations from the average structure do exist. Therefore, we calculated a 3D variance map (16) (right column in Fig. 2). Whereas no significant differences are observed in the CP, a number of conspicuous variances are seen in the RPs. One particularly prominent variance hotspot is visible in the RP near the interface between the base and the lid (see below).

**Structural Analysis of the ATPase Density.** Little is currently known about the subunit topology of the RPs, beyond the assignment of Rpt1–6, Rpn1–2, and Rpn10 to the proximal (base) part and Rpn3–12 to the distal (lid) part (17). Interaction patterns between subunits, as derived from yeast 2-hybrid screens (18) or
excellently into the 6-fold-symmetrized EM map. We did not attempt to model restraints are the yeast 2-hybrid interactions between Rpt4 and Rpt6 as well as the present single-particle reconstruction show 6 distinct centers of mass at the base of the RP mass (see Fig. S1), consistent with a (pseudo)-6-fold symmetry. Therefore, we have in a first step applied 6-fold symmetry to the entire complex around its long axis (Fig. 3A). The 2 RPs were then separated from the CP by identifying the depicted minima of a projection plot (see Fig. 3A) of the reconstruction. By assuming a protein density of 1.3 g/cm³ and a molecular mass of 300 kDa for the ATPase hexamer, a subvolume was extracted from the reconstructed volume. Finally, the exact position and orientation of the ATPase subvolume was refined by an exhaustive 6D (3 translational and 3 rotational parameters) search procedure. The search was iterated until no significant changes of the position and orientation parameters of the ATPase subvolume were found (Fig. 3B).

The resulting structure has a hexagonal base, measuring 12 nm across the flats, and on top of it a “mouth” with a diameter of 7.9 nm; the total height is 7.2 nm. The segmented structure is highly reminiscent of the structure of the PAN complex, the archetypal activator of the 20S proteasome that is found in some archaea (20). Likewise reassuring is the fact that in our Drosophila data set there is a subpopulation of approximately 900 particles that upon averaging is almost congruent with the 20S CP plus the segmented ATPase, probably representing (dis-)assembly intermediates of the 26S holocomplex (Fig. S2). There is an additional rather elongate mass embracing the ATPase module and making contact to α-subunits of the CP; we tentatively assigned this mass to subunits Rpn1 and Rpn2 (Fig. S3). At the current resolution it is not possible to delineate the boundaries of Rpn1 or Rpn2 throughout the map. Our assignment is based on the reported physical interactions of Rpn1 and Rpn2 with the AAA-ATPase (21) and with the α-ring of the 20S CP (22), as well as on their predicted domain architecture: Rpn1 and Rpn2 likely consist of relatively small HEAT repeats (38–50 residues each) enabling the formation of extended structures (23).

The most intriguing finding is that the (pseudo-) 6-fold symmetry axis of both ATPase modules does not coincide with the (pseudo-) 7-fold symmetry axis of the CP; both ATP modules are shifted by ~3.0 nm with respect to the axis of the CP. This finding indicates that the assembly mechanism of 20S/CP is markedly different from the complex of CP and 11S activator; the 11S particle is 7-fold symmetric and its symmetry axis coincides with the one of the CP in the 20S/11S holocomplex (24). For the 20S/AAA-ATPase the symmetry mismatch does not allow the ATPase subunits to interact with the subunits of the α-ring in an equivalent manner. In fact, only 3 of the ATPase subunits contain a motif at the C terminus that can insert into pockets in the α-subunits and induce gate opening (25). In addition to the different symmetry of the AAA-ATPases compared with the 11S activator, the requirement of ATP for the assembly of the 19S holocomplex in contrast to the 20S/11S complex makes it plausible that the architectures of these holocomplexes are not similar.

Interestingly, the axis of one of the 2 ATPase modules in the double-capped complexes is almost inclined with respect to the 7-fold axis of the core complex, breaking the 2-fold symmetry of the holocomplex (Fig. 3C). We assume that the different inclinations at both cylinder ends reflect 2 different functional states of the CP. Our experimental observation is in excellent agreement with the “wobbling model” as proposed in ref. 26, which is based on theoretical considerations. In this model, an inclination of the ATPase relative to the 20S CP is implicated in the mechanism of opening the 20S CP gate.

Each subunit consists of a non-ATPase N-domain and an AAA-fold (27); each N-domain is further predicted to consist of a 25- to 60-residue N-terminal region, a coiled-coil, and an intermediate segment. The N-domain of the Archaeoglobus fulgidus ortholog PAN (PAN-N) could be crystallized as a homohexamer (41). We built comparative models of the N-domains and the AAA-domains of Rpt1–6 by using PAN-N and the AAA-folds of FtsH (PDB code 2ce7) and p97 (1e32) as structural templates. Next, we determined the configuration, that is, the order of Rpt1–6 in the hexamer by using these models.
as well as published biochemical interactions as determined by yeast 2-hybrid assays, chemical cross-linking, filter binding, copurification, and different pulldown experiments (Table S1). We built all possible arrangements of the ATPases and scored them on the basis of the proteomics data (see Materials and Methods). None of the models satisfied all experimental restraints, yet 12 of 120 possible configurations violated only 2 restraints of 16. Among these models, only 1 configuration, Rpt1/2/6/3/4/5, allowed an invariant proline in Rpt2, -3, and -5 (corresponding to Pro-62 of PAN-N), to adopt a cis isofrom, which is required for coiled-coil formation. Notably, this configuration differs from the previously proposed order Rpt1/2/6/4/5/3 (28), which violates 3 restraints. The difference between our configuration and the previously determined one can be attributed to interactions reported in the meantime. In addition, the narrow pore of the N-domain ring (~11 Å) can be spanned by the cross-linkers that were used (28), which was presumably not anticipated (Fig. 3D). After establishing the order of the ATPases we rebuilt homology models for N- and AAA-hexamers and fitted them into the EM map (Fig. 3E). Using the EM map, we can thus obtain a model for the structure of N-domain and AAA-domain. An ambiguity remains because a (anti-clockwise) rotation of the N-domain by 60° results in a conformation that would also be compatible with the linker length.

RPN1-RPN2 Are Not Localized in the Center of the ATPase. Recently, structural and biochemical studies performed with different mixtures of purified Rpn1, Rpn2, and 20S CP but in the absence of the ATPase module led to the proposal that Rpn1/Rpn2 may form a heterodimer of 2 stacked concentric rings, which is physically attached to the 20S CP (22). For the complete 26S proteasome, this model further implies that the Rpn2/Rpn1 dimer is localized in the center of the AAA-ATPases and might provide a direct physical link between substrate recruitment and proteolysis. However, the internal volume of our AAA-ATPase hexamer model could accommodate a maximum of ~75 kDa, whereas the putative Rpn1/Rpn2 dimer has a molecular mass of ~220 kDa. Moreover our EM map shows that the cavity enclosed by the ATPase harbors no significant protein mass. Therefore, our data on the 26S holocomplex do not support the model proposed for 20S/Rpn2/Rpn1 (22).

Proposed Position of RPN10. The variance map (Fig. 2) indicates that a substantial mass may be present or absent in 1 of the 2 RPs. To elaborate on this observation, we performed a 3D maximum-likelihood classification (ML3D) on these data (29). The question arises as to the identity of the variable mass. The fact that it is found only in ~50% of all particles analyzed and that it is present in only 1 of the 2 RPs implies that it is present in a substoichiometric amount of approximately 1:4. The only subunit found in a similar stoichiometry is Rpn10 (see Fig. 1). Rpn10 has a molecular mass of 42.6 kDa, which is in good agreement with the mass (60 ± 25 kDa) derived from the volume of the additional mass in the class averages, when using the mass of the 20S CP for calibration. Although the C-terminal ubiquitin-interacting motif of Rpn10 (approximately 1/3 of the sequence) was shown to be disordered in solution (30), most of the density of Rpn10 should show up in a low-resolution map. We also tried to confirm the position of Rpn10 by assembly modeling (31). In a preliminary approach, we represented all proteins within the 26S proteasome as strings of spheres, each representing a single protein domain, and built models that comply with the EM data and reported physical protein–protein interactions (Fig. S4). Although this approach did not yield a unique solution, it is notable that the proposed position of Rpn10 coincides with 1 of 3 predominant locations of Rpn10 in the models.

Taken together, our observations provide strong circumstantial evidence, albeit not a rigorous proof, that the variable mass is Rpn10. Interestingly, there is a whole body of experimental data suggesting that Rpn10 is a receptor for mult ubiquitin chains and can bind to the 26S proteasome reversibly (32). Its location close to the mouth of the ATPase module appears to be well placed to deliver substrates for the ATP-dependent processing steps. At present, no high-resolution structure of Rpn10 is available that could be used for docking into the EM density. Given the fact that the C-terminal half of Rpn10 was shown to be highly flexible in solution, it might be difficult to obtain such a structure (30). However, it might be possible to locate Rpn10 by using ubiquiinated substrate bound to it and to take snapshots of early steps of substrate uptake and processing by the RPs.

Materials and Methods

Purification and Mass Spectrometry. 26S proteasomes from Drosophila were prepared as described in refs. 11 and 12. Fractions used for structural studies were loaded on a native gradient (2–15%) polyacrylamide gel and subjected to electrophoresis, and proteasome bands were excised and analyzed by mass spectrometry using a hybrid linear ion trap–orbitrap instrument (Thermo Fisher Scientific Inc.) (33). Although not as accurate as isotope-based technologies (34), the “label-free” proteomics approach used in this study achieves good quantitation results for high-intensity signals (35). The measured intensities for the 26S proteasome sample were found to be in the high-range region (109 to 1010 counts per second) where accuracy of the measured signals is high. We determined the stoichiometry of the sample by normalizing the sum of all of the identified peptides by the molecular masses of the individual proteins. The resulting intensities were standardized relative to the subunits of the 20S CP; only small variations (standard deviation of 0.24) were found, indicating that the procedure yields accurate enough results.

EM. EM data were recorded in a fully automated manner using a microscope equipped with an energy filter that was operated in a 0-loss mode at an accelerating voltage of 300 kV. Magnification on the detector plane (CCD
camera) was 82.500%, corresponding to a 0.36 nm pixel size in the object plane. The defocus was set to nominal values of 2 μm and 4 μm, respectively; at these settings, the first 0 of the contrast transfer function (CTF) is at 2.0 nm and 2.8 nm. Micrographs and their corresponding power spectra were visually inspected. Altogether, 5,134 micrographs of a total of 11,605 were selected for further processing. The contrast transfer functions were determined and outliers were set to a weighted average of successfully determined defocus values of the surrounding micrographs (for details see ref. 14). The selected electron micrographs were deconvoluted by a phase-flipping procedure described elsewhere (36). Particles showing the typical shape of the 26S proteasome holocomplex (205 core flanked by two 195 complexes) were selected in an interactive manner. Two particle stacks with a box size of 160 × 160 pixels were created, comprising 18,931 (2 μm underfocus) and 17,285 (4 μm underfocus) particles.

**3D Variance Analysis and Classification.** The first 3D reconstruction was low-pass-filtered to a resolution of 6 nm and used to generate 3 seeds for subsequent ML3D classification that yielded 1 class with a strongly enhanced additional mass. Both classes were refined separately by progressively reducing the angular sampling interval to 5° and finally 2°. Final reconstructions were obtained by using an ART with blobs algorithm (37) and the data from the 2-μm and 4-μm defocus subsets were combined by using an 3D Wiener filter (38), resulting in a resolution of 2.0 nm at these settings, the first 0 of the contrast transfer function (CTF) is at 2.0 nm.

**AAA-ATPase Modeling.** First, we built comparative models of the N-domains and the AAA-domains of Rpt–1 by using PAN-H and the AAA-ffolds of FshD (PDB code 2ce7) and p97 (1e32) as structural templates using MODELLER (39). For the initial model we assumed the previously proposed order of the ATPases Rpt1/2/6/4/5/3 (28). Both rings were then fitted separately into the 6-fold-symmetrized EM map by using Chimera (40), whereby the rotation of the N-ring with respect to the AAA-ring was chosen such that domains of the same protein are adjacent.

We further elucidated the order of the Rpt–1 subunits in the ring by assessing 51 AAA-ring configurations (the position of Rpt1 was kept fixed) against reported interactions found in the literature (Table S1). Specifically, each reported interaction between a subset of subunits was compiled into a conditional connectivity restraint (31) that we solved using the current minimal distance between the corresponding Cα atoms by using a harmonic function. The expected minimal distance between interacting subunits is a function of errors of the comparative model; in case the interaction restraint is derived from chemical cross-links, we added the length of the linker arm to the expected minimal distance.

**ACKNOWLEDGMENTS.** We thank Kornelius Zeth and Andrei Lupas for insightful discussions and for providing us with the atomic models of PAN-N. This work was supported by the 3-D Repertoire Grant, the 3-D EM Network of Excellence grant, and the High throughput-3DEM grant, all within the Research Framework Program 6 (FP6) of the European Commission. A.S. was supported by the National Institutes of Health (R01 GM54762, RO1 GM083960, and US4 RR022220), National Science Foundation (IIS-0705196), Hewlett-Packard, IBM, Intel, and NetApp. S.H.W.S. and J.M.C. were supported by the National Institutes of Health (HL070472) and by the Spanish Comision Interministerial de Ciencia y Tecnologia (BFU2004-02017), the Ministerio de Educacion y Ciencia (CSO2006-00023, BIO2007-67150-C03-1 and -3), and Comunidad Autonoma de Madrid S-GEN-0166-2006. We acknowledge the Barcelona Supercomputing Center (Centro Nacional de Supercomputacion) for providing computing resources. K.L. has been funded by the Clere Foundation Predoctoral Fellowship. F.F. is grateful for a long-term fellowship from the Human Frontier Science Program Organization.