Hidden dynamics of human proteasome autoregulation during protein degradation

Zhaolong Wu[1,2], Shuwen Zhang[1], Wei Li Wang[1,3], Yuanchen Dong[4], Youdong Mao[1,2,3]*

[1]State Key Laboratory for Artificial Microstructure and Mesoscopic Physics, School of Physics, Peking University, Beijing, China. [2]Center for Quantitative Biology, Peking University, Beijing, China. [3]Dana-Farber Cancer Institute, Boston, MA, USA. [4]Institute of Chemistry, Chinese Academy of Sciences, Beijing, China.

* To whom correspondence may be addressed. Email: ymao@pku.edu.cn

Abstract

The 2.5-MDa 26S proteasome maintains proteostasis and regulates myriad cellular processes. How polyubiquitylated substrate interactions regulate proteasome activity is not understood. Here we reconstitute hidden dynamics of proteasome autoregulation during protein degradation by cryo-electron microscopy. We identified 64 conformers of the substrate-bound human 26S proteasome via deep classification. These structures unexpectedly revealed substrate-dependent conformational entanglement of two regulatory particles in the doubly capped holoenzymes and energetic differences between doubly and singly capped ones. Novel ubiquitin-binding sites were delineated on the RPN2, RPN10 and α5 subunits, which remodel polyubiquitin chains for deubiquitylation and recycling. Importantly, single-nucleotide-exchange dynamics of proteasomal AAA-ATPase motor were choregraphed during translocation initiation, which upregulates proteolytic activity by allosterically promoting nucleophilic attack. Our systematic analysis illuminates a grand hierarchical allostery for proteasome autoregulation.

The 26S proteasome is the known largest ATP-dependent protease machinery of ~2.5 MDa molecular weight ubiquitously found in all eukaryotic cells. It exists at the center of the ubiquitin-proteasome system (UPS) that regulates myriad cellular processes, such as protein quality control, cell cycle, gene expression, inflammation and immunity [1,2]. Dysregulation of the UPS has been broadly associated with carcinogenesis, ageing, neurodegenerative and cardiovascular diseases [1]. The 26S holoenzyme is assembled by a cylindrical 20S core particle (CP) capped with one or two 19S regulatory particles (RPs), each consisting of the lid and base subcomplexes [1,2]. The ring-like heterohexameric motor of ATPases-associated-with-diverse-cellular-activities (AAA) ATPase in the base subcomplex mechanically unfold polyubiquitylated substrates recognized by ubiquitin receptors in the lid subcomplex. Previous structural data by cryogenic electron microscopy (cryo-EM) have suggested three modes of coordinated ATP hydrolysis in the AAA-ATPase motor that regulate distinct steps of ubiquitin recognition, deubiquitylation, initiation of translocation and processive substrate translocation in the proteasome [2]. However, key intermediate conformations between these states were missing, greatly limiting our understanding of the intrinsic proteasome regulation. The mechanism of the proteasome targeting of substrate-conjugated ubiquitin signals is not fully understood [1,3-5]. How polyubiquitylated substrate interactions regulate proteasome activity remains enigmatic [1].

The 26S proteasome is one of the most complex, dynamic, conformationally and compositionally heterogeneous holoenzyme machinery in cells [1,2,6-9], challenging the current approaches in cryo-EM structure determination [3,10,11]. Visualizing atomic structures of transient, nonequilibrium intermediates connecting the major states of the proteasome has been unsuccessful by conventional cryo-EM analysis [1,2,9]. To address this major challenge, we recently developed a deep manifold learning framework that can break such a limitation and enable atomic-level cryo-EM reconstructions of highly dynamic, lowly populated intermediates or transient states [12]. By applying this approach to analyze a large cryo-EM dataset that previously solved the atomic structures of 7 proteasomal RP-CP states [2], we reconstructed 64 conformers of the human 26S proteasome in the act of substrate degradation up to 2.5-Å resolution. Systematic analyses of those transient or intermediate states in substrate processing pinpoint energetic differences between the doubly and singly capped proteasomes, discover several novel ubiquitin-binding sites in the proteasome, and choreograph single-nucleotide-exchange dynamics of the proteasomal AAA-ATPase motor during translocation initiation. Importantly, the dramatically enriched proteasomal states provide unprecedented insights into an

integrative, hierarchical allosteric mechanism of proteasome autoregulation upon polyubiquitylated substrate interactions.

Visualizing hidden dynamics of the proteasome

We conducted deep 3D classification on a large cryo-EM dataset of the substrate-engaged 26S proteasome (see Materials and Methods; table S1) [2]. This dataset was collected on the human 26S proteasome mixed with a Lys63-linked polyubiquitinated substrate Sic1^{PY} for 30 seconds before ATP was diluted with ATPγS, which was expected to maximally capture any intermediate states before the degradation reaction was completed [2]. The reconstruction of the energy landscape of the substrate-engaged proteasome suggests that the rate-limiting step of substrate degradation lies in the initiation of substrate unfolding after substrate deubiquitylation, reflected in the highest activation energy barrier between states E_B and E_{C1} (Fig. 1, B and C), consistent with a recent fluorescence microscopy study on the proteasome kinetics [13]. Comparison with the energy landscape of the substrate-free ATPγS-bound 26S proteasome [8] indicates that substrate interactions considerably expand the conformational space of the proteasome toward the states of open CP gates (Fig. 1A).

Our deep 3D classification allowed reconstructions of significantly more conformers of the proteasomes in various compositional forms from the same cryo-EM dataset that previously yielded seven atomic structures of the proteasome [2]. It discovered 20 conformational states of RP-CP subcomplex (Fig. 1, B and C; figs. S2 and S3; and table S1). Nine transient states (designated $E_{A1.1}$, $E_{A1.2}$, $E_{A2.2}$, $E_{A2.3}$, $E_{B.2}$, $E_{B.3}$, $E_{D2.2}$, $E_{D2.3}$ and $E_{D.\alpha5}$) at 3.1-7.5 Å resolution exhibit previously unseen features resembling ubiquitin that map various putative ubiquitin-binding sites on the RPN1, RPN2, RPN10 and α5 subunits in the proteasome (Fig. 2, fig. S1, F to K, and movie S1). Six sequential intermediate states (designated $E_{D0.1}$, $E_{D0.2}$, $E_{D0.3}$, $E_{D1.1}$, $E_{D1.2}$ and $E_{D1.3}$) at 3.2-3.9 Å resolution bridge a major missing gap between states E_{C2} and E_{D1}, revealing how concerted structural transitions in the AAA-ATPase motor during single-nucleotide exchange drive the initiation of substrate unfolding (Figs. 3 and 4, movie S2).

To understand whether the singly capped (SC) and doubly capped (DC) proteasomes behave differently, we reconstructed 8 and 36 conformational states for the pure SC and DC holoenzymes, respectively (Fig. 1, D and E, figs. S4 and S5). All 8 SC and 29 DC states were refined to 3.7-4.7 Å, with 7 DC states limited to 6-9 Å due to their transient nature and extremely low population (fig. S4, B, E, and F). Note that the CP gate in the RP-distal side of the 20 states

of RP-CP subcomplex is in a heterogeneous conformation of various stoichiometric ratios between closed and open states. By contrast, both CP gates in the 44 states of pure SC or DC reconstructions are in a homogeneous conformation.

Besides resolving significantly more conformers, improved 3D classification pushes the envelope of the achievable resolution of known states by keeping more particles without sacrificing conformational homogeneity. For example, the local resolution of RPN11-bound ubiquitin in state $E_{A2.1}$ was improved from 5-6 Å to 3.5 Å (fig. S3A). It reveals a previously unsolved feature of a short substrate stretch bound to the cleft between RPN11 and RPT4 at the ATPase entrance prior to substrate insertion into the ATPase ring in state E_B (fig. S6). The structure of state E_{D2} was also improved from previous resolution at 3.2 Å to 2.5 Å, allowing us to investigate how the CP gating allosterically regulates the catalytic activity of the β-type subunits (Fig. 5).

Conformational entanglement

Both the SC and DC proteasomes are abundant in cells [14]. Their structural and functional difference, however, remains elusive. In the 36 DC reconstructions, each conformer is a combination of two RP states. Each CP gate in the DC proteasome appears to be rigorously controlled by their respective proximal RP and is only open when its proximal RP is in an E_D-compatible state with five RPT C-terminal tails inserted into the surface pockets on the α-ring [2] (Fig. 1E and fig. S4, A to D). Consistently, the gate on the RP-free side of the CP in the SC proteasome is closed in all states, indicating that the CP gate opening on the RP-controlled side does not remotely open the other CP gate on the RP-free side (fig. S4D).

Previous studies have found no obvious coupling of two RP conformations in the same DC proteasome in the absence of substrate [8,14]. To understand whether this is also true in the presence of substrate, we compare the state distribution matrix obtained by experimental data to the prediction of a control model assuming two completely uncoupled RP states in the DC proteasome (Fig. 1, F to H). Although the overall pattern of the experimental distribution roughly agrees with the model prediction, a notable number of DC states exhibit substantial deviation, indicating that the conformational dynamics of the two RPs in the same proteasome are profoundly entangled together (Fig. 1, G and H). Intriguingly, the most prominent positive deviations are seen in the state distributions of E_{A1}-E_{A1}, E_B-E_B, E_B-E_{D2}, and E_{D2}-E_{D2} combinations (Fig. 1G), suggesting that the deubiquitylation state E_B and the translocation state

E_{D2} of one RP are allosterically coupled to and mutually upregulated by states E_B and E_{D2} of the opposite RP. Coupled actions of substrate deubiquitylation and translocation between two opposite RPs are expected to optimize the efficiency of substrate processing. Moreover, all the DC states along the diagonal of the state distribution matrix are more or less upregulated (indicated by red in Fig. 1, G and H), suggesting that both RPs in the same proteasome synergistically promote each other to transit into downstream states in a symmetrical fashion.

Notably, the largest discrepancy of the distribution of RP states between the SC and DC proteasomes also lies in state E_B (Fig. 1, B to D). The state population of E_B is significantly lower in the SC than in the DC proteasome. This observation verifies the conclusion drawn from the above analysis on the DC proteasomes alone that state E_B is upregulated by the conformational entanglement effect in the DC holoenzyme, suggesting that the SC proteasome lacks the functional synergy between two opposite RPs seen in the DC proteasome (Fig. 1H). Altogether, these results implicate that the DC proteasome could be energetically more efficient in substrate deubiquitylation and degradation than the SC proteasome.

Ubiquitin-proteasome interactions

Unexpectedly, in states $E_{A1.2}$, $E_{A2.2}$, and $E_{B.2}$, a diubiquitin-like density is consistently found on RPN2, whereas a monoubiquitin-like density is observed at the apex of the von Willebrand factor type A (VWA) domain [15] of RPN10 in states $E_{A2.3}$, $E_{B.3}$ and $E_{D2.3}$ (Fig. 2A and fig. S3, B to F). In state $E_{D2.2}$, the diubiquitin-like density, however, appears to contact both RPN2 and RPN10, and can be modelled with a Lys63-linked diubiquitin (fig. S3, B and G). The RPN10 ubiquitin-interacting motifs (UIMs), however, were not reliably observed presumably due to their highly flexible nature (fig. S3G). The populations of these states are very small and represented in only a few thousand particles or 0.1-0.3% of the entire dataset, indicating their transient nature. The local resolutions of the RPN2- and RPN10-bound ubiquitin-like densities are lower than the rest of the reconstructions and other ubiquitin densities on RPN1, RPN11 and α5 subunits. Rigid-body fitting of the ubiquitin structure into these ubiquitin-like densities confirms the agreement of the overall density shape with the ubiquitin fold (Fig. 2A and fig. S3, B to G).

RPN1 sites. By focused 3D classification, we refined the density map of RPN1 in state $E_{A1.0}$ (~5.9% of the dataset) to 3.7 Å, which was previously reconstructed at moderate resolution (~5.5 Å) [2] (figs. S1D and S3D). It reveals that a putative ubiquitin moiety contacts Glu458, Cys459

and Asp460 at the N-terminal end of an RPN1 helix (homologous to H24 between the T1 and T2 sites of yeast Rpn1) and potentially interacts with the T2 site (Asp423, Leu426 and Asp430) on the RPN1 toroidal domain (Fig. 2, C and D) [5]. Because the yeast Rpn1 T2 site was found to preferentially bind Ubp6 ubiquitin-like (UBL) domain [5], this site may be sterically occluded for ubiquitin binding in the presence of USP14/Ubp6. In addition, we also identified another state $E_{A1.1}$ that exhibits a diubiquitin-like density with one ubiquitin moiety on the RPN1 T1 site and the other near the T2 site (fig. S3E). The much smaller population of state $E_{A1.1}$ (~1.3 % of the dataset) is consistent with the previous finding that Lys63-linked polyubiquitin chains have a lower affinity against the T1 site than those linked via Lys48 and Lys6 residues [5].

RPN2 site. The ubiquitin-binding site on RPN2 is centered on three helices harboring an acidic surface around residues Tyr502, Glu530, Asp531 and Ser569 (Fig. 2, B to E). Structural comparison suggests that the ubiquitin-binding site on RPN2 is homologous to the T1 site of RPN1 (Fig. 2, C and D) [5]. We therefore name it 'the RPN2 T1 site'. Although there is little conservation in amino acid sequences between the RPN1 and RPN2 T1 sites, their structural and electrostatic properties are highly conserved (Fig. 2D, fig. S3, H to K). Comparison with the NMR structure of Lys48-linked diubiquitin-bound RPN1 T1 site helices [5] further confirms the conservation and similarity of the ubiquitin-binding modes between the RPN1 and RPN2 T1 sites (Fig. 2C). Consistent with this finding, previous studies have observed weak interactions of UBL proteins and RPN2 by cross-linking experiments [16]. As there is also ubiquitin bound to RPN11 in states $E_{A2.2}$, and $E_{B.2}$, a Lys63-linked triubiquitin chain can be fitted into the RPN2- and RPN11-bound ubiquitin densities by modeling the rotation of Lys63 linkages, suggesting that the RPN2 T1 site may function as a coreceptor site for ubiquitin transfer from RPN10 to RPN11 to facilitate deubiquitylation (Fig. 2G). Given that at least one RPN10 UIM-bound ubiquitin invisible in these cryo-EM maps would be expected to exist, the observation of triubiquitin on RPN2 and RPN11 structurally rationalizes why tetraubiquitin is an optimal degradation signal for the proteasome [1].

RPN10 VWA site. In states $E_{A2.3}$, $E_{B.3}$ and $E_{D2.3}$, the monoubiquitin-like densities consistently reside on Lys132, Lys133 and Lys163 of the RPN10 VWA domain. By contrast, in state $E_{D2.2}$, a diubiquitin-like density appears to span from this lysine-rich site of RPN10 to the RPN2 T1 site and can be modelled with a Lys63-linked diubiquitin (fig. S3, B and G). In line with these observations, previous studies have suggested that the yeast Rpn10 VWA is a ubiquitin-binding domain and the ubiquitylation of Rpn10 VWA itself prevents Rpn10 from being incorporated

into the proteasome [15]. As ubiquitylation of RPN10 was found to regulate substrate recruitment [15,17-19], it remains to be clarified whether the putative ubiquitin at this lysine-rich site is covalently modified by ubiquitylation of RPN10 itself co-purified from of HEK293 cells.

α5-subunit site. In state $E_{D2.2}$, we can observe a ubiquitin density bound to the RPN1 T2 site and another weaker ubiquitin density at the α5-subunit ~6 nm beneath the RPN1 T2 site (Fig. 2A, fig. S7B). To understand if the ubiquitin-binding site on the α5 subunit is used more broadly, we combined all the E_D-compatible states and focused the 3D classification on the α5-bound ubiquitin. This yielded a much larger 3D class, named state $E_{D.\alpha5}$ (~4% of the entire dataset) that improved the ubiquitin-bound CP structure to 3.1 Å, and the α5-bound ubiquitin density to 4.9 Å (figs. S1, E, H, F and K, S3F). The ubiquitin-binding site on the α5 subunit is composed of acidic or hydrophobic residues Glu183, Val184, His186, Ser188 and Glu193 in a loop connecting two helices, highly resembling the mode of ubiquitin recognition by an RPN13 loop[4,20] (Fig. 2F, fig. S3F). The electrostatic property of this site is also similar to those of the RPN1 T1/T2 and RPN2 T1 sites that are all acidic, in charge complementarity with the basic ubiquitin surface around residues Ile44 and His68 (fig. S3, H to O). Although no previous evidence supports ubiquitin interactions with the CP, recent data suggest that the free 20S CP can degrade substrate-conjugated ubiquitin, which might be facilitated by ubiquitin binding to the α-ring [21].

We note that the putative ubiquitin-binding sites reside in the vicinity of primary receptor sites. The RPN2 T1 site is very close to the RPN10 UIMs, whereas the α5 site is near the RPN1 T2 site. Modelling of a Lys63-linked tetraubiquitin chain by fitting its terminal ubiquitin into the cryo-EM density suggests that a single tetraubiquitin can span across the α5 subunit and the RPN1 T2 site (Fig. 2H). Given that RPN1 appears to allosterically regulate the AAA-ATPase motor conformation [2] (Fig. 3, A and B), tetraubiquitin binding to both the RPN1 T2 site and α5 subunit could allosterically stabilize the open-CP states (E_{D0}, E_{D1} and E_{D2}). It is unclear whether these low-affinity ubiquitin-binding sites would be directly involved in primary targeting of substrate-conjugated polyubiquitin chains by the proteasome in cells. However, it is conceivable that the polyubiquitin chains could be first captured by the high-affinity sites and then moved to the nearby low-affinity sites. Working together, these ubiquitin-binding sites could remodel polyubiquitin chains and deliver the peptide-proximal ubiquitin to the deubiquitylation site [2] (Fig. 2G) as well as assist ubiquitin recycling after isopeptide bond hydrolysis (Fig. 2H). Due to the limited ubiquitin resolution and transient nature of these ubiquitin-bound states, we cannot rule out that some of the ubiquitin-like densities are co-purified endogenous ubiquitin, UBL

proteins or ubiquitylation of the proteasome itself [1,22]. Future functional studies are required to understand their definitive chemical nature and biological importance in cells.

Single-nucleotide exchange dynamics

The six intermediate states ($E_{D0.1}$, $E_{D0.2}$, $E_{D0.3}$, $E_{D1.1}$, $E_{D1.2}$ and $E_{D1.3}$) during initiation of substrate unfolding and translocation exhibit highly coordinated movements in the RP relative to the CP (Fig. 3, A to C, fig. S7), with prominent rotations in the RPN1 N-terminal that detaches from RPT2 in states $E_{D0.1}$, $E_{D0.2}$, and $E_{D0.3}$ and re-binds RPT2 in state $E_{D1.1}$, $E_{D1.2}$ and $E_{D1.3}$. All six conformations share the same architecture in their substrate interactions with pore-1 loops from four RPT subunits (RPT2, RPT6, RPT3 and RPT4), which remain largely unchanged in overall conformation (Fig. 3D). These structures illustrate that the RPT2 re-association with the substrate proceeds that of RPT1 after both RPT1 and RPT2 are dissociated from the substrate in state E_C. They also share the same pattern of nucleotide states, with ATP bound to RPT1, RPT2, RPT6 and RPT3 and ADP bound to RPT4. There are poor, partial nucleotide densities in the RPT5 nucleotide-binding pocket, which give rise to the highest B-factor (250-500) when fitted with ADP, suggesting that RPT5 undergoes ADP release in these states (fig. S7K).

Markedly, the pore-1 loop of RPT1 gradually moves toward the substrate from a distance of ~16 Å to ~3 Å, whereas the pore-1 loop of RPT5 gradually moves away from the substrate from a distance of ~3 Å to ~16 Å (Fig. 4A). The small AAA subdomain of RPT5 moves inward together with the large AAA subdomain of RPT1, as the large AAA subdomain of RPT5 rotates away from the substate (fig. S7F-I). Consistently, the arginine fingers (R-finger) of RPT2 are progressively moved toward ATP bound in RPT1, closing up the nucleotide-binding pocket (Fig. 3E and 4B). In contrast, the nucleotide-binding pocket of RPT5 is gradually shrunk but is not fully closed by the R-fingers of RPT1, whereas the nucleotide-binding pocket of RPT4 is gradually opened up (Fig. 4B). As a result, the pore-1 loops of RPT1 and RPT5 are only in contact with the substate in the last and first states $E_{D1.3}$ and $E_{D0.1}$, respectively. Both are dissociated from the substate in states $E_{D0.2}$, $E_{D0.3}$, $E_{D1.1}$, and $E_{D1.2}$. These gradual movements allow us to assign their temporal sequences as illustrated (Fig. 3, A to E), indicating that these states represent intermediate conformations accompanying single-nucleotide exchange that occurs in RPT5 (Fig. 4C). Interestingly, the outward flipping of RPT5 appears to complete its major conformational change in state $E_{D0.3}$, much faster than the inward motion of RPT1, which completes its conformational change in state $E_{D1.3}$ (Fig. 4A). This observation suggests that

ATPase re-engagement with the substrate is the rate-limiting step in the single-nucleotide exchange kinetics of the AAA-ATPase motor.

Our data clarify that the CP gate is open once the pore-1 loop of RPT2 binds the substrate in state $E_{D0.1}$, as five RPT C-tails are inserted in the α-pockets in all six states (fig. S4C). Altogether, these structures vividly characterize at the atomic level how the "hand-over-hand" actions of RPT subunits are executed by coordinated conformational dynamics in the AAA-ATPase motor during single-nucleotide exchange and generates mechanical force unfolding the substrate. Although the sequential model of coordinated ATP hydrolysis around the AAA-ATPase ring has been hypothesized in numerous similar systems [1,2,6,9], our results provide the first direct evidence for sequential intermediate states of hand-over-hand actions within a single cycle of nucleotide exchange.

Hierarchical allosteric regulation of proteolytic activity

Both atomic structures of the substrate-bound CPs in the open- and closed-gate states E_A and E_{D2} show a segment of substrate polypeptide binding the catalytically active residue Thr1 in the β2 subunit, whereas no substrate density is observed in the other two catalytic subunits β1 and β5. In both states, the substrate is bound in a β-sheet conformation between residues 20 to 22 and 46 to 48 in the β2 subunit (Fig. 5, A and B, fig. S8, A and B), reminiscent of the conformations of CP inhibitors binding the β5 subunits in the crystal structures of the CP [23,24]. The hydroxyl oxygen atom in the catalytic Thr1 is ~2-Å closer to the substrate mainchain in state E_{D2} compared to state E_A. This structural difference is expected to result from the CP gate opening that exerts a profound allosteric impact on the entire CP structure [25-28].

To understand how such a subtle structural change influences the proteolytic activity, we carried out first-principles calculations of the quantum-mechanical interactions involved in the nucleophilic attack of the catalytic residue Thr1 on the scissile peptide bond in the substrate, using the density functional theory (DFT) with the generalized gradient approximation [29]. To initiate the nucleophilic attack, the carbonyl group needs to be brought to the proximity of the electrophile Thr1-Oγ to facilitate formation of a tetrahedral transition state [24,30,31]. We calculated the charge density difference induced around the interacting pair after proton transfer from Thr1-Oγ to Thr1-NH_2 (Fig. 5, C and D, fig. S8, E and F). The substrate-interacting pair of atoms in state E_{D2} form a bonding region of slight positive charge difference linking the carbonyl group of the substrate to Thr1-Oγ. By contrast, the pair are separated by a charge depletion non-bonding

region in state E_A. Indeed, a recent study has found that the nucleophilic attack is the rate-limiting step in the acylation barrier [30]. Our result suggests that CP gate opening allosterically promotes the nucleophilic attack by lowering the activation energy barrier of acylation and induces hybridization of molecular orbitals, hence upregulating the reaction rate of proteolysis. In support of our structural finding, previous studies have observed indirect evidence for coupling of the CP gating with the proteolytic activity [25-28].

Taken together, our systematic structural analyses point to a grand allostery model for the proteasomal autoregulation of proteolytic activity through a three-tier hierarchy of inter-molecular interaction networks (Fig. 6), in which all subunit conformations are allosterically coupled. The first tier of autoregulation is between the CP gate in the α-ring and the catalytic sites of the β-subunits. Opening of the CP gate allosterically upregulates the proteolytic activity of the three β-subunits. The second tier of autoregulation lies between the substrate-bound AAA-ATPase motor and the rest of the proteasome. The AAA-ATPase motor bridges the ubiquitin-binding sites and the CP via direct interactions of the RPT coiled-coil (CC) domains with RPN1, RPN2 and RPN10 harboring ubiquitin-binding sites. As a result, three modes of coordinated ATP hydrolysis appear to regulate the intermediate steps of substrate processing [1,2], in which initiation of substrate unfolding is allosterically coupled with CP gating during E_{C2} to $E_{D0.1}$ transition (Fig. 3). The third tier of autoregulation is via dynamic interactions between ubiquitin signals and a collection of ubiquitin receptor sites and deubiquitinases. This is manifested as the dual roles of ubiquitin-binding sites in recognizing ubiquitin signals and in transmitting such recognitions to allosteric regulation of AAA-ATPase conformations that control the CP gate. Ubiquitin binding and remodeling destabilizes the resting state of the proteasome and promotes its transition to states E_B, E_C and ultimately E_D. On the other hand, a polyubiquitin chain binding multiple sites in the proteasome could have profound allosteric and conformational selection effects (Fig. 1, G and H). Such three-tier, hierarchical allosteric regulations may occur mutually with both RPs and are expected to underlie the conformational entanglement of the two RPs in the DC proteasome (Fig. 1E), which was not clearly observed in the absence of substrates [1,7,8]. This grand allostery mechanism might have been evolved to optimize the proteasome function in response to extremely complicated intracellular challenges and is manifested as various phenotypic effects, such as UBL-induced proteasome activation [32,33] and long-range allosteric couplings in the proteasome [1,26,27,34,35].

References and Notes

1 Mao, Y. Structure, Dynamics and Function of the 26S Proteasome. In: Harris J.R., Marles-Wright J. (eds) Macromolecular Protein Complexes III: Structure and Function. *Subcell Biochem* **96**, 1-151, doi:10.1007/978-3-030-58971-4_1 (2021).

2 Dong, Y. *et al.* Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49-55, doi:10.1038/s41586-018-0736-4 (2019).

3 Galves, M., Rathi, R., Prag, G. & Ashkenazi, A. Ubiquitin Signaling and Degradation of Aggregate-Prone Proteins. *Trends Biochem Sci*, doi:10.1016/j.tibs.2019.04.007 (2019).

4 Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481-488, doi:10.1038/nature06926 (2008).

5 Shi, Y. *et al.* Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351**, doi:10.1126/science.aad9421 (2016).

6 Zhang, S. & Mao, Y. AAA+ ATPases in Protein Degradation: Structures, Functions and Mechanisms. *Biomolecules* **10**, doi:10.3390/biom10040629 (2020).

7 Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proc Natl Acad Sci U S A* **113**, 12991-12996, doi:10.1073/pnas.1614614113 (2016).

8 Zhu, Y. *et al.* Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. *Nat Commun* **9**, 1360, doi:10.1038/s41467-018-03785-w (2018).

9 de la Pena, A. H., Goodall, E. A., Gates, S. N., Lander, G. C. & Martin, A. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **362**, doi:10.1126/science.aav0725 (2018).

10 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).

11 Scheres, S. H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157, doi:10.1016/bs.mie.2016.04.012 (2016).

12 Wu, Z. *et al.* Visualizing conformational space of functional biomolecular complexes by deep manifold learning. *bioRxiv*, doi:10.1101/2021.08.09.455739 (2021).

13 Bard, J. A. M., Bashore, C., Dong, K. C. & Martin, A. The 26S Proteasome Utilizes a Kinetic Gateway to Prioritize Substrate Degradation. *Cell* **177**, 286-298 e215, doi:10.1016/j.cell.2019.02.031 (2019).

14 Asano, S. *et al.* Proteasomes. A molecular census of 26S proteasomes in intact neurons. *Science* **347**, 439-442, doi:10.1126/science.1261197 (2015).

15 Keren-Kaplan, T. *et al.* Structure of ubiquitylated-Rpn10 provides insight into its autoregulation mechanism. *Nat Commun* **7**, 12960, doi:10.1038/ncomms12960 (2016).

16 Saeki, Y., Sone, T., Toh-e, A. & Yokosawa, H. Identification of ubiquitin-like protein-binding subunits of the 26S proteasome. *Biochem Biophys Res Commun* **296**, 813-819, doi:10.1016/s0006-291x(02)02002-8 (2002).

17 Uchiki, T. *et al.* The ubiquitin-interacting motif protein, S5a, is ubiquitinated by all types of ubiquitin ligases by a mechanism different from typical substrate recognition. *J Biol Chem* **284**, 12622-12632, doi:10.1074/jbc.M900556200 (2009).

18 Isasa, M. *et al.* Monoubiquitination of RPN10 regulates substrate recruitment to the proteasome. *Mol Cell* **38**, 733-745, doi:10.1016/j.molcel.2010.05.001 (2010).

19 Lipinszki, Z., Kovacs, L., Deak, P. & Udvardy, A. Ubiquitylation of Drosophila p54/Rpn10/S5a regulates its interaction with the UBA-UBL polyubiquitin receptors. *Biochemistry* **51**, 2461-2470, doi:10.1021/bi3001006 (2012).

20 Schreiner, P. *et al.* Ubiquitin docking at the proteasome through a novel pleckstrin-

homology domain interaction. *Nature* **453**, 548-552, doi:10.1038/nature06924 (2008).
21 Sahu, I. *et al.* Signature activities of 20S proteasome include degradation of the ubiquitin-tag with the protein under hypoxia. *bioRxiv*, 2019.2012.2020.883942, doi:10.1101/2019.12.20.883942 (2019).
22 Besche, H. C. *et al.* Autoubiquitination of the 26S proteasome on Rpn13 regulates breakdown of ubiquitin conjugates. *EMBO J* **33**, 1159-1176, doi:10.1002/embj.201386906 (2014).
23 Schrader, J. *et al.* The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science* **353**, 594-598, doi:10.1126/science.aaf8993 (2016).
24 Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* **386**, 463-471, doi:10.1038/386463a0 (1997).
25 Osmulski, P. A., Hochstrasser, M. & Gaczynska, M. A tetrahedral transition state at the active sites of the 20S proteasome is coupled to opening of the alpha-ring channel. *Structure* **17**, 1137-1147, doi:10.1016/j.str.2009.06.011 (2009).
26 Haselbach, D. *et al.* Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat Commun* **8**, 15578, doi:10.1038/ncomms15578 (2017).
27 Kleijnen, M. F. *et al.* Stability of the proteasome can be regulated allosterically through engagement of its proteolytic active sites. *Nat Struct Mol Biol* **14**, 1180-1188, doi:10.1038/nsmb1335 (2007).
28 Ruschak, A. M. & Kay, L. E. Proteasome allostery as a population shift between interchanging conformers. *Proc Natl Acad Sci U S A* **109**, E3454-3462, doi:10.1073/pnas.1213640109 (2012).
29 Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys Rev Lett* **77**, 3865-3868, doi:10.1103/PhysRevLett.77.3865 (1996).
30 Saha, A., Oanca, G., Mondal, D. & Warshel, A. Exploring the Proteolysis Mechanism of the Proteasomes. *J Phys Chem B* **124**, 5626-5635, doi:10.1021/acs.jpcb.0c04435 (2020).
31 Huber, E. M. *et al.* A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat Commun* **7**, 10900, doi:10.1038/ncomms10900 (2016).
32 Collins, G. A. & Goldberg, A. L. Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation. *Proc Natl Acad Sci U S A* **117**, 4664-4674, doi:10.1073/pnas.1915534117 (2020).
33 Kim, H. T. & Goldberg, A. L. UBL domain of Usp14 and other proteins stimulates proteasome activities and protein degradation in cells. *Proc Natl Acad Sci U S A* **115**, E11642-E11650, doi:10.1073/pnas.1808731115 (2018).
34 Rennella, E., Huang, R., Yu, Z. & Kay, L. E. Exploring long-range cooperativity in the 20S proteasome core particle from Thermoplasma acidophilum using methyl-TROSY-based NMR. *Proc Natl Acad Sci U S A* **117**, 5298-5309, doi:10.1073/pnas.1920770117 (2020).
35 Yu, Z. *et al.* Allosteric coupling between alpha-rings of the 20S proteasome. *Nat Commun* **11**, 4580, doi:10.1038/s41467-020-18415-7 (2020).
36 Wang, X. *et al.* Mass spectrometric characterization of the affinity-purified human 26S proteasome complex. *Biochemistry* **46**, 3553-3565, doi:10.1021/bi061994u (2007).
37 Verma, R., McDonald, H., Yates, J. R., 3rd & Deshaies, R. J. Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase cyclin-Cdk. *Mol Cell* **8**, 439-448 (2001).
38 Verma, R., Oania, R., Graumann, J. & Deshaies, R. J. Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. *Cell* **118**, 99-

110, doi:10.1016/j.cell.2004.06.014 (2004).
39 Saeki, Y., Isono, E. & Toh, E. A. Preparation of ubiquitinated substrates by the PY motif-insertion method for monitoring 26S proteasome activity. *Methods Enzymol* **399**, 215-227, doi:10.1016/S0076-6879(05)99014-9 (2005).
40 Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60, doi:10.1016/j.jsb.2005.03.010 (2005).
41 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).
42 Wu, J. *et al.* Massively parallel unsupervised single-particle cryo-EM data clustering via statistical manifold learning. *PLoS One* **12**, e0182130, doi:10.1371/journal.pone.0182130 (2017).
43 Unverdorben, P. *et al.* Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc Natl Acad Sci U S A* **111**, 5544-5549, doi:10.1073/pnas.1403409111 (2014).
44 Wehmer, M. *et al.* Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc Natl Acad Sci U S A* **114**, 1305-1310, doi:10.1073/pnas.1621129114 (2017).
45 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
46 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
47 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
48 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).
49 The PyMOL Molecular Graphics System Version 2.0 (Schrödinger, LLC).
50 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).
51 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65, doi:10.1038/nmeth.2727 (2014).
52 Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* **157**, 3-18, doi:10.1016/j.jsb.2006.06.006 (2007).
53 Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665-667, doi:10.1093/nar/gkh381 (2004).
54 Soler, J. M. *et al.* The SIESTA method for ab initio order-N materials simulation. *J Phys Condens Matter* **14**, 2745-2779 (2002).

Acknowledgments: The authors thank A. Goldberg, D. Finley, S. Elsasser, M. Kirschner, Y. Lu, and for constructive discussions. The cryo-EM data were collected at the Cryo-EM Core Facility Platform and Laboratory of Electron Microscopy at Peking University. The data processing was performed in the High-Performance Computing Platform at Peking University.

Funding: Beijing Natural Science Foundation grant Z180016/Z18J008 (YM). National Natural Science Foundation of China grant 11774012 (YM)

Author contributions: Y.M. conceived this study. S.Z., Y.D., Z.W. and W.L.W. contributed to the experimental cryo-EM data acquisition. Z.W. and S.Z. analyzed the experimental cryo-EM datasets and refined the density maps. E.C. and W.L.W. conducted the quantum mechanical simulation. Z.W. and Y.M. verified the density maps, built and refined the atomic models and analyzed the data. Y.M. supervised this study and wrote the manuscript with inputs from all authors.

Competing interests: The authors declare no competing financial interests.

Data and materials availability: The atomic coordinates and their corresponding cryo-EM density have been deposited into the RCSB Protein Data Bank (PDB) (www.rcsb.org) and Electron Microscopy Data Bank (EMDB) (www.emdatasource.org). Raw cryo-EM data have been deposited into the Electron Microscopy Pilot Image Archive (www.ebi.ac.uk/pdbe/emdb/empiar).

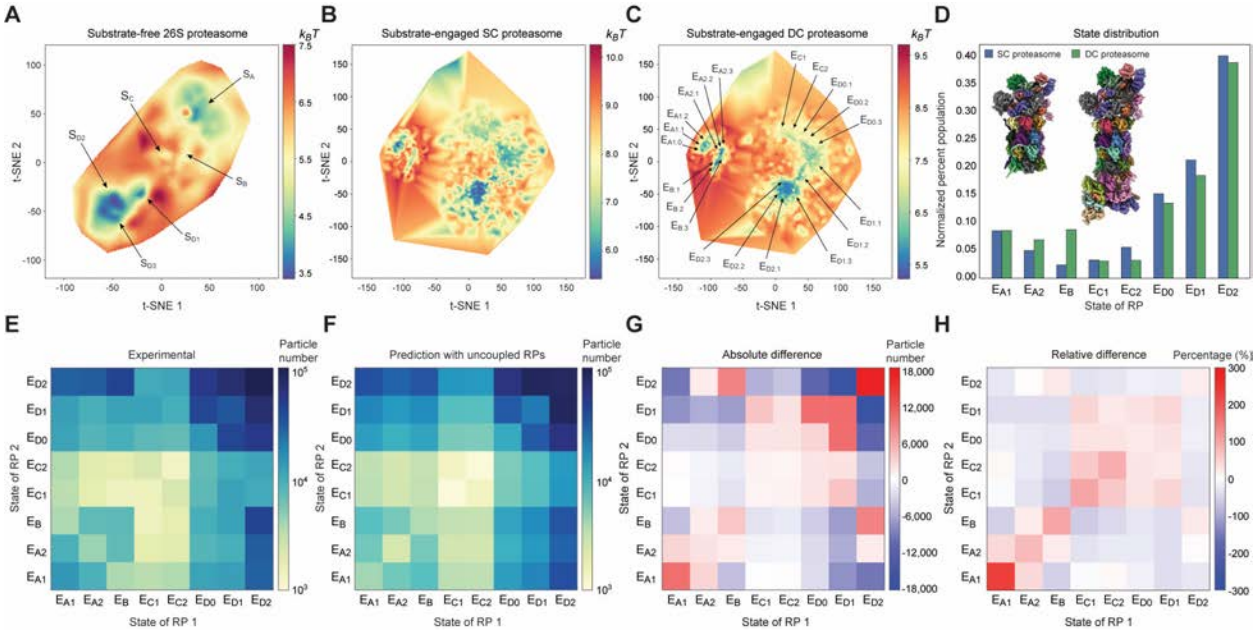

Fig. 1. Energetic differences of the singly and doubly capped human proteasomes. (**A**-**C**) Energy landscape of the substrate-free ATPγS-bound 26S proteasome (**A**), the substrate-engaged SC proteasome (**B**), and the RP-CP subcomplex of the substrate-engaged DC proteasome (**C**). The energy landscape of the substrate-free proteasome shows six local energy minima corresponding to the known six conformational states S_A, S_B, S_C, S_{D1}, S_{D2} and S_{D3} [8]. In contrast, 20 conformational states of the RP-CP subcomplex are marked on the energy landscape of the substrate-engaged 26S proteasome. (**D**) Comparison of the state distribution of RP in the SC and DC proteasomes. (**E**) State distribution matrix of the DC proteasome obtained using experimental data, colored by the particle numbers in the DC states, with the horizonal and vertical axes representing the states of two RPs bound to the same CP. (**F**) State distribution matrix of the DC proteasome predicted by a control model assuming the two completely uncoupled RPs in the same DC proteasome. The model assumes the same total number of DC particles and the same probability of observing each RP state as experimentally measured. (**G**) State distribution difference matrix of the DC proteasome by experimental results minus the model predictions, colored by the absolute particle number difference. (**H**) Relative differences of the DC proteasome computed as the ratio of absolute difference divided by the particle numbers from the model prediction in each DC state.

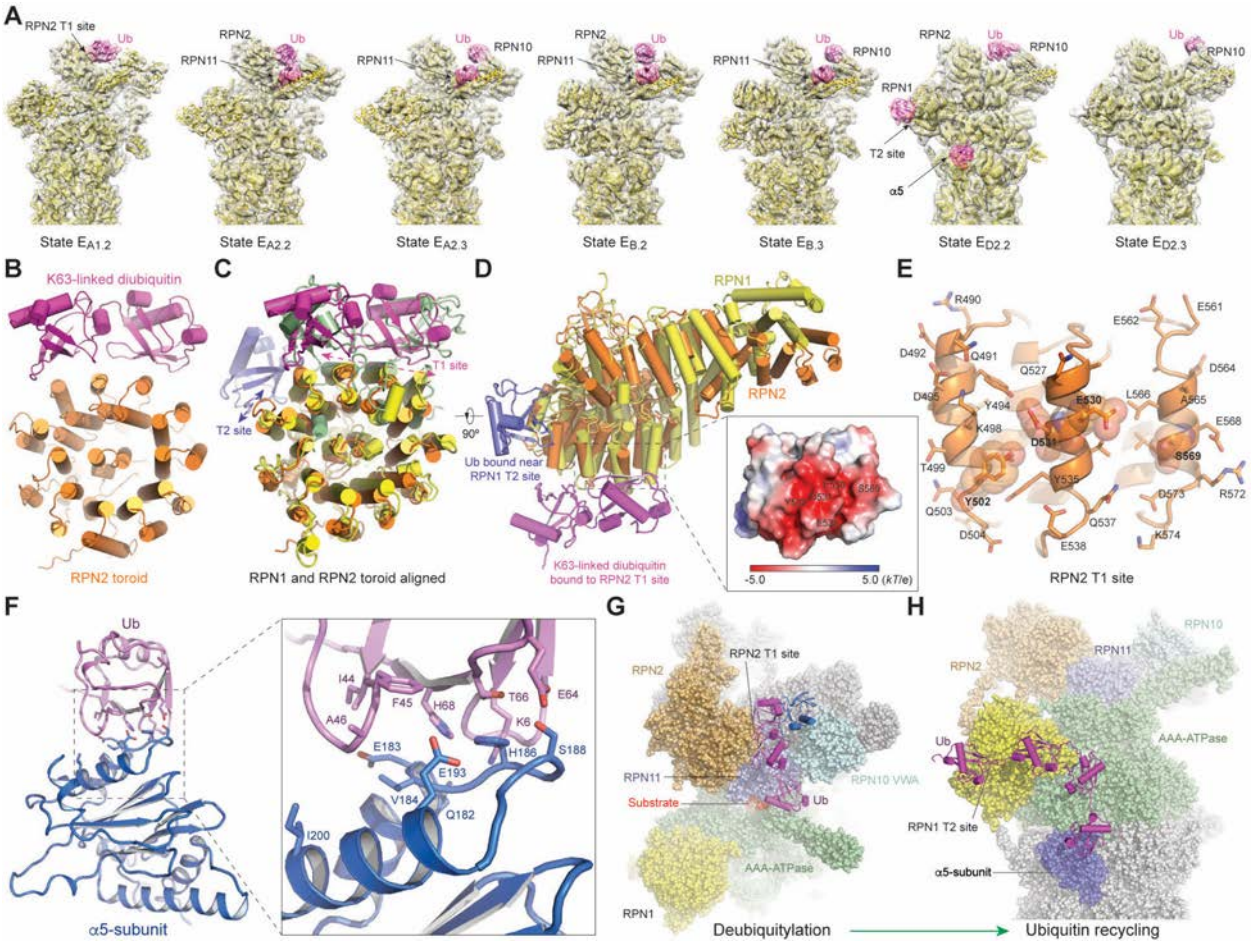

Fig. 2. Mapping of ubiquitin-binding sites on the human 26S proteasome. (**A**) Seven lowly populated states visualize various ubiquitin-binding sites on the RPN1, RPN2, RPN10 and α5 subunits. For clarity, the cryo-EM maps are low-pass filtered at 8 Å and shown in transparent surface representations superimposed with their corresponding atomic models in cartoon representations. The ubiquitin (Ub) densities are colored magenta. (**B**) Structure of the RPN2 toroidal domain (orange) in complex with Lys63-linked diubiquitin (magenta) modelled from the cryo-EM maps of states $E_{A1.2}$, $E_{A2.2}$ and $E_{B.2}$, in cartoon representation. (**C**) Structure-based alignment of the human RPN1 (yellow) and RPN2 (orange) toroidal domains. Ubiquitin bound near the RPN1 T2 site is shown as light blue. The NMR structure (PDB ID 2N3V) of the Lys48-linked dibuiquitin-bound yeast Rpn1 T1 site segment [5], shown in light green, is aligned and superimposed with the human RPN1 and RPN2. (**D**) The aligned human RPN1 and RPN2 are viewed from a perspective orthogonal to that in panel (**C**). Insert, electrostatic surface of the RPN2 T1 site showing its acidic nature. (**E**) Closeup view of the RPN2 T1 site composed of three helices in carton representation with the side chains shown in stick representation. The key residues contacting ubiquitin are highlighted with transparent sphere representation. (**F**) Structure of ubiquitin-bound α5 subunit shown in cartoon representation. The side chains of residues mediating intermolecular interactions are shown in stick representation. (**G**) A Lys63-linked triubiquitin chain modelled from the cryo-EM maps of states $E_{A2.2}$ and $E_{B.2}$ is shown in purple cartoon representation and spans from the RPN2 T1 site to the RPN11 deubiquitylation site. The RPN10 VWA-bound monoubiquitin fitted in states $E_{A2.3}$ and $E_{B.3}$ is superimposed in blue cartoon representation. The proteasome is shown as space-filling sphere representations.

(**H**) A Lys63-linked tetraubiquitin chain model derived from the cryo-EM map of state $E_{D2.2}$. The two terminal ubiquitin molecules are fitted from their highest-resolution density maps available at the RPN1 T2 and α5-subunit sites, whereas the middle two molecules are hypothetically modelled.

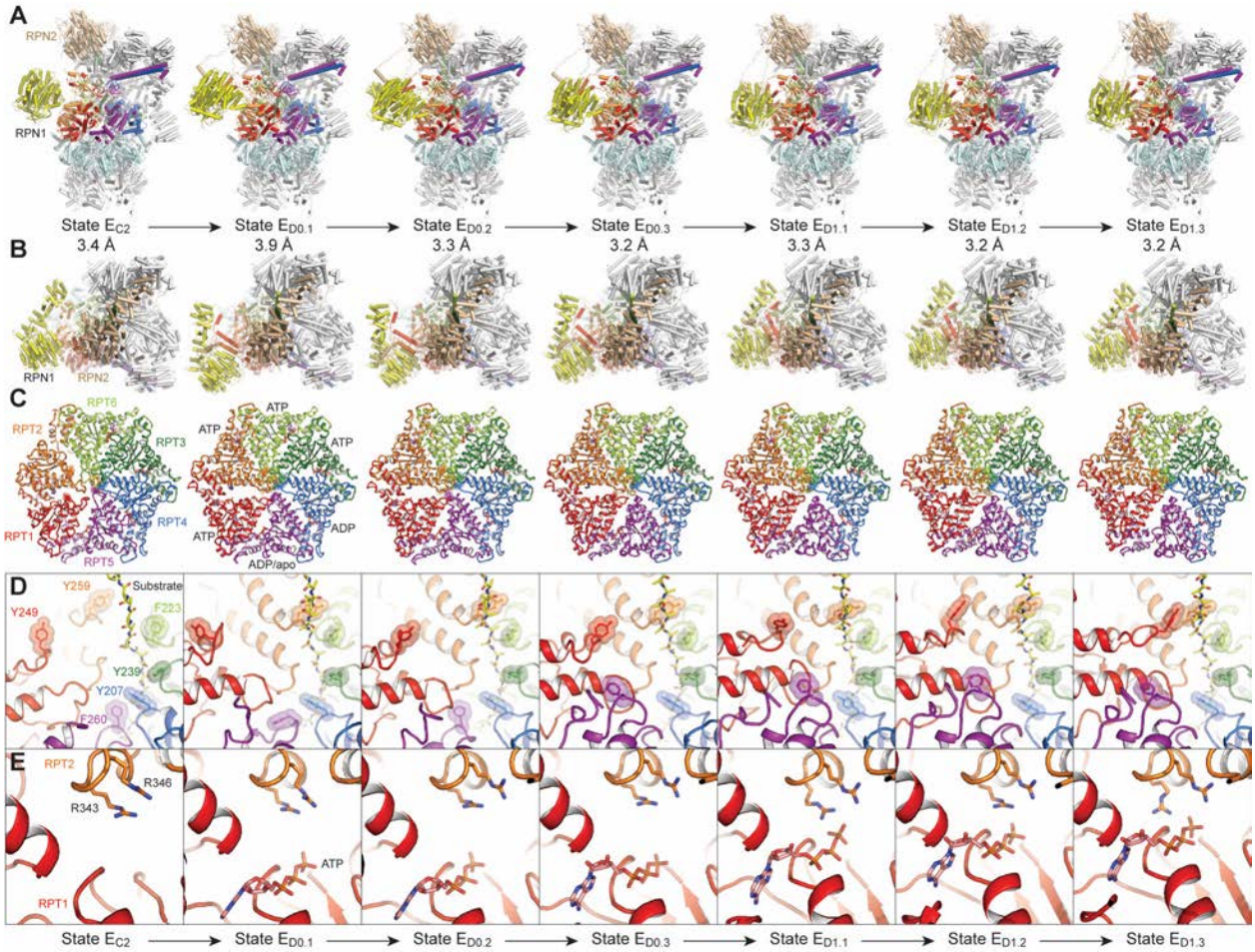

Fig. 3. Novel intermediate states of the substrate-engaged 26S proteasome during single-nucleotide exchange. (**A**) Side views of the substate-engaged RP-CP subcomplex of seven sequential conformers in cartoon representation. (**B**) Top views of the seven sequential conformers, showing the overall lid rotation and the relative rotation between RPN1 and the rest of the RP structure. (**C**) The AAA-ATPase structures of the seven conformers in cartoon representations. The substrates and nucleotides are shown in stick representations. (**D**) Close-up views of the pore-1 loop interactions with the substrate (in yellow stick representation), with the aromatic residues in the pore-1 loops highlighted by transparent sphere representations. (**E**) Close-up views of the RPT1 nucleotide-binding sites in different states.

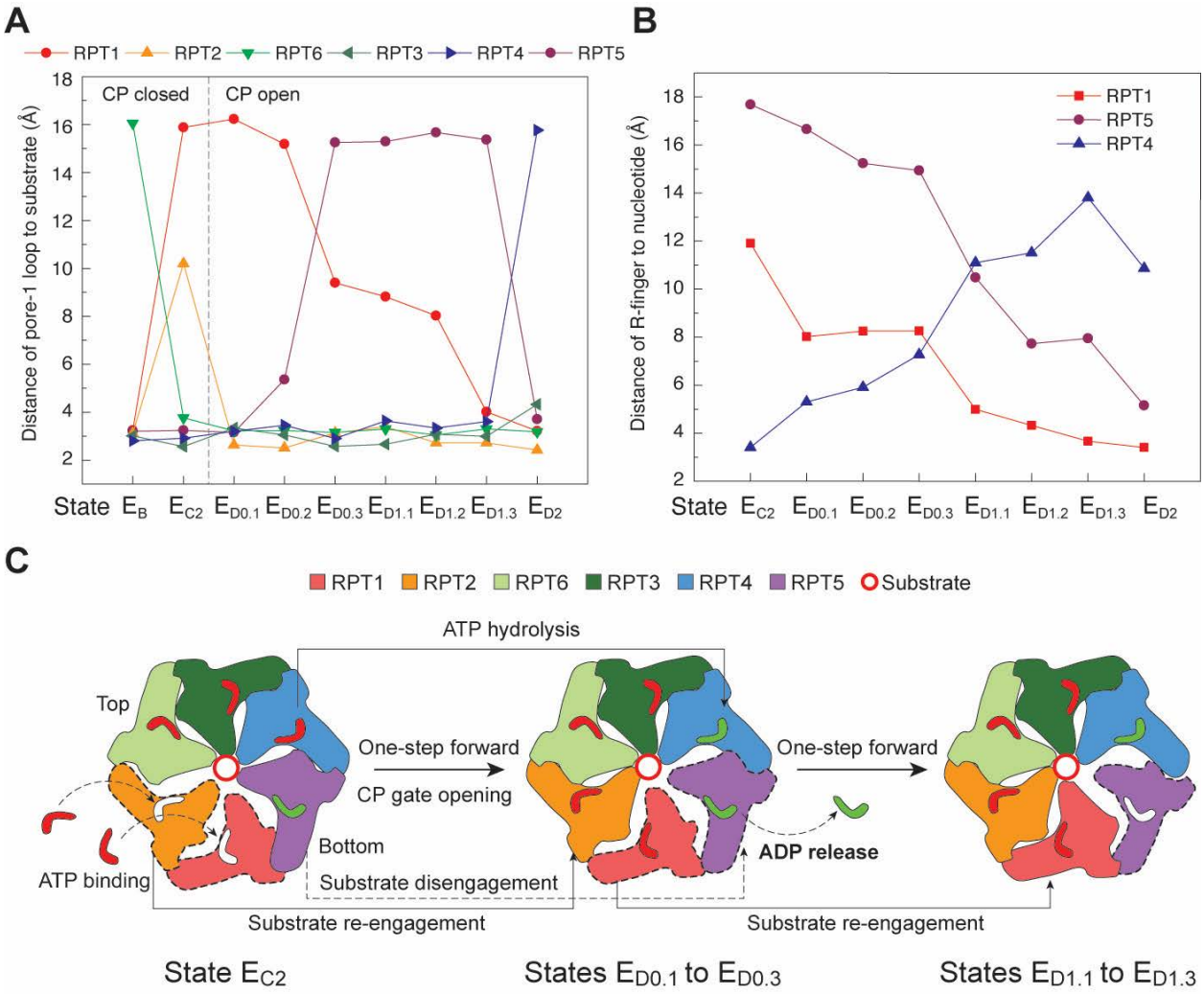

Fig. 4. Single-nucleotide-exchange dynamics of the proteasomal AAA-ATPase motor during translocation initiation. (**A**) Plots of distances of the pore-1 loops of all RPT subunits to the substrate measured in the atomic structures of 9 sequential conformational states. (**B**) Plots of distances of R-finger to the nucleotides bound to three RPT subunits measured in the atomic structures of 8 sequential conformational states. (**C**) Schematic illustrating the coordinated ATP hydrolysis reaction and nucleotide exchange associated with the intermediates during the transition from state E_{C2} to $E_{D1.3}$.

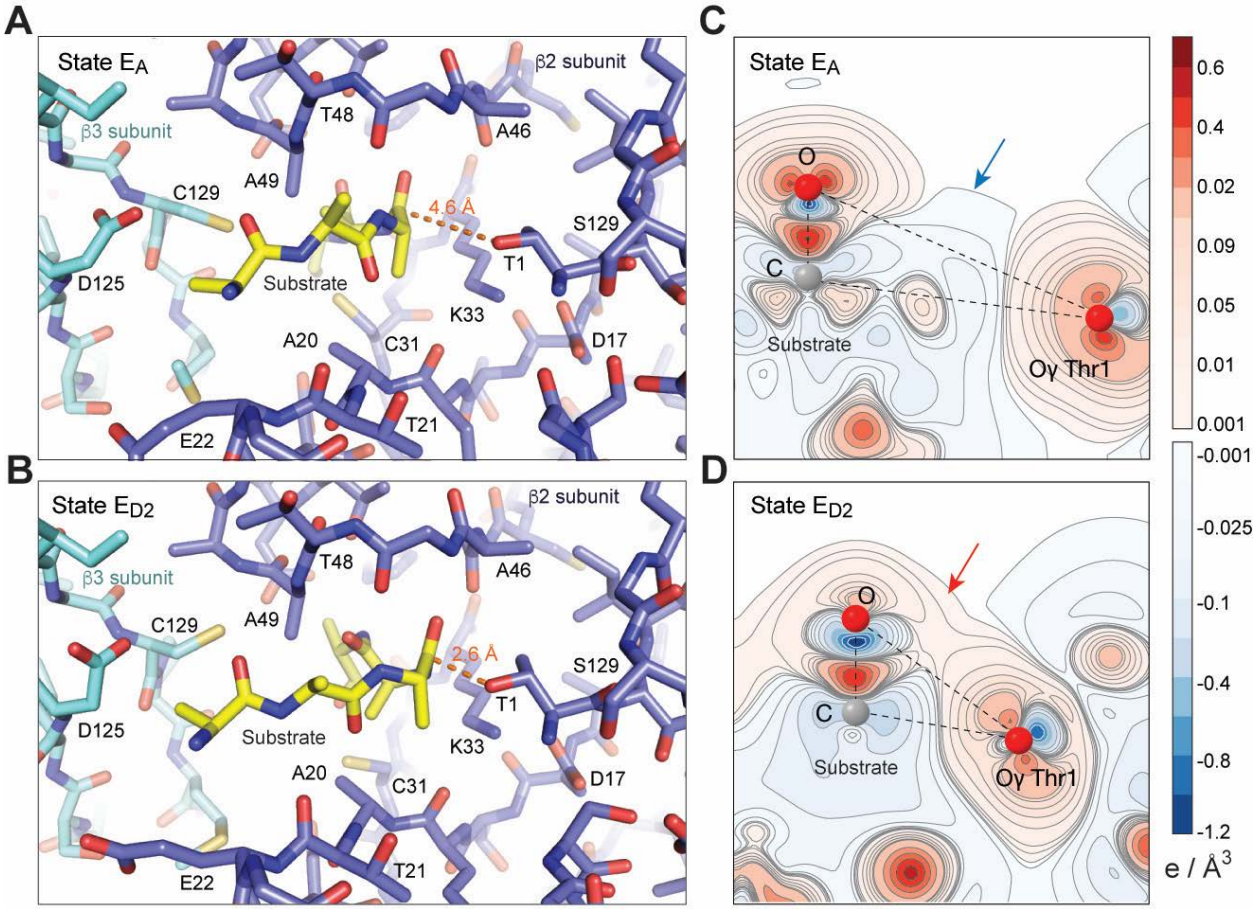

Fig. 5. Allosteric regulation of substrate proteolysis by the opening of CP gate. (**A** and **B**), Structure of the catalytic site with a substrate polypeptide bound at residue Thr1 in the β2 subunit of the CP in state E_A at 2.7-Å resolution (**A**) and in state E_{D2} at 2.5-Å resolution (**B**). c and d, Charge density difference contour maps at the proteolytic active site of the β2 subunit computed by the quantum mechanical DFT based on the atomic structures of states E_A (**C**) and E_{D2} (**D**). The contours are plotted in the plane passing through the Thr1-Oγ atom and the nearest carbonyl carbon and oxygen atoms of the substrate. Blue and red arrows highlight a non-bonding charge depletion region in state E_A and a bonding interaction indicated by the slight positive charge difference region in state E_{D2}, respectively.

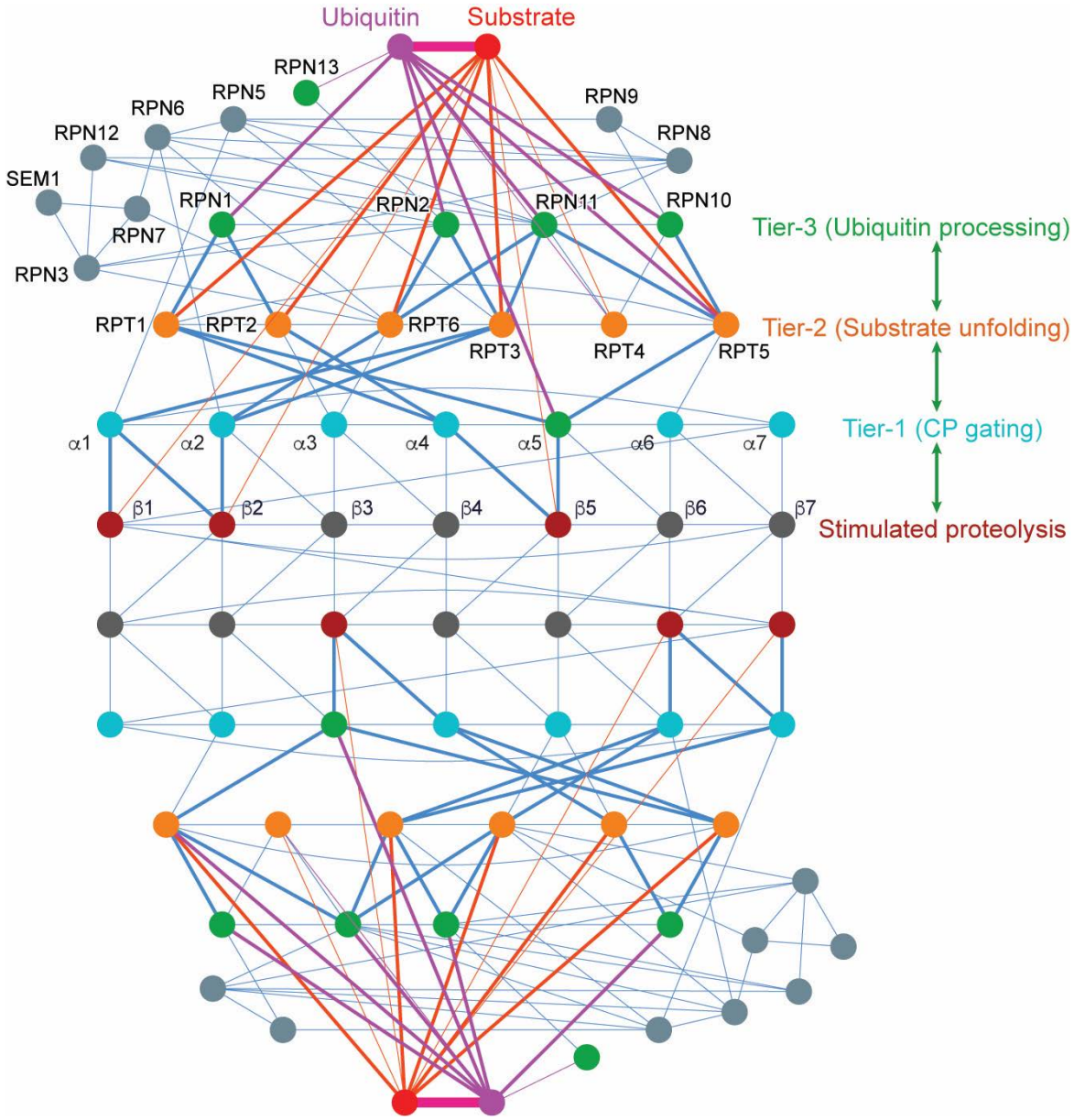

Fig. 6. A network model of hierarchical allosteric regulation of proteasome activity. The interactome network diagram illustrating the grand allostery of the proteasome. Steel blue lines represent inter-subunit interactions; purple lines represent the interactions with ubiquitin; and red lines represent substrate polypeptide interactions. Bold lines highlight the shortest pathways connecting the three tiers of regulatory subunits that are expected to propagate the allosteric regulation from ubiquitin interactions down to the six proteolytic active sites in the CP. The network diagram is constructed according to all 64 structural models of the substrate-bound proteasome in this study. Notably, the subnetwork architecture of RPN1 connecting to the proteolytic subunits is identical to that of RPN2 connecting to the proteolytic subunits, consistent with the structural homology of RPN1 and RPN2 as well as their potential functional similarity in ubiquitin interactions.

Materials and Methods

Protein purification and sample preparation

Human 26S proteasomes were affinity-purified as previously described [2]. In brevity, a stable HEK293 cell line expressing HTBH (hexahistidine, TEV cleavage site, biotin, and hexahistidine)-tagged hRPN11 (a gift from L. Huang, University of California, Irvine)[36] was used to purify the human proteasome on a large scale. After Dounce-homogenization in a lysis buffer (50 mM PBS [pH 7.5], 10% glycerol, 5 mM $MgCl_2$, 0.5% NP-40, 5 mM ATP, 1 mM DTT, and protease inhibitors), cell lysates were cleared by centrifugation (20,000 × g) for 30 min. The supernatant was incubated with NeutrAvidin agarose resin (Thermo Scientific) for 3 h at 4 °C. 26S proteasomes were cleaved from the agarose beads by TEV protease (Invitrogen) after washing with excess lysis buffer and a wash buffer (50 mM Tris-HCl [pH 7.5], 1 mM $MgCl_2$ and 1 mM ATP). The supernatant was then further purified by gel filtration on a Superose 6 10/300 GL column at a flow rate of 0.15 ml/minute in buffer (30 mM Hepes [pH 7.5], 60 mM NaCl, 1 mM $MgCl_2$, 10% glycerol, 0.5 mM DTT, 0.6 mM ATP). Gel-filtration fractions were concentrated to about 2 mg/ml and the buffer was exchanged to 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM ATP and 10% glycerol.

The model substrate $Sic1^{PY}$ was expressed in BL21 (DE3) *Escherichia coli* cells and purified as previously described [2]. The PY motif, Pro-Pro-Pro-Tyr, is recognized by WW domains the Rsp5 family of ligases. In the $Sic1^{PY}$ construct, a PY motif was inserted to the N-terminal segment (MTPSTPPSRGTRYLA) of the Cdk inhibitor Sic1 [37,38], resulting in a modified N-terminus of MTPSTPP<u>PPY</u>SRGTRYLA. The ubiquitylation reaction of $Sic1^{PY}$ was conducted by mixing 40 µg/ml $Sic1^{PY}$ with 500 nM E1 Ube1 (Boston Biochem), 2 µM E2 UbcH5a (Boston Biochem) and 100 µg/ml E3 WWHECT and 1 mg/ml ubiquitin (Boston Biochem) in a reaction buffer (50 mM Tris-HCl [pH 7.5],100 mM NaCl, and 10% glycerol, 2 mM ATP, 10 mM $MgCl_2$, and 1 mM DTT) for 3 h at room temperature. The recombinant E3 ligase WWHECT was engineered from deletion of Rsp5 N-terminal 220 amino acids, which conjugates ubiquitin chains via Lys63 linkage [39]. Pre-equilibrated TALON resin was then incubated with the sample for 1 h at 4°C. After the resin was washed with 20 column volumes of the wash buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10% glycerol), the polyubiquitylated $Sic1^{PY}$ was eluted with the same buffer containing 150 mM imidazole. The elution was applied to an Amicon ultrafiltration device with 30K molecular cut-off for removal of imidazole. The ubiquitylation results were examined by Western blotting with anti-T7 antibody.

Cryo-EM data collection

The purified proteasome (10 µl of 2 mg/ml) was mixed with polyubiquitylated Sic1^{PY} (9 µl of 2 mg/ml; molar ratio ~3:1 for substrate vs proteasome) in a buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10% Glycerol and 1 mM ATP) at room temperature. After 30 s of incubation of the mixture, 1 µl 20 mM ATPγS was immediately added into the solution to quench the degradation reaction. To remove the glycerol, the complex system was applied to Zeba™ Micro Spin Desalting Columns (7K, Thermo Fisher Scientific, MA, USA), exchanging the buffer to 50 mM Tris-HCl (pH 7.5) containing 100 mM NaCl, 1mM ATP and 1 mM ATPγS. The glycerol removal process usually took about 10 min before cryo-plunging. 0.005% NP-40 was added to the proteasome solution immediately prior to cryo-plunging. A 2.5-µl drop of 1.5 mg/ml substrate-engaged proteasome solution was applied to the glow-discharged C-flat grids (R1/1 and R1.2/1.3, 400 Mesh, Protochips, CA, USA) in a Vitrobot Mark IV (Thermo Fisher Scientific, MA, USA) chamber with 100% humidity at 4 °C. After 2 s of blotting, the grid was plunged into liquid ethane and then transferred to liquid nitrogen. Cryo-EM images were collected on an FEI Titan Krios G2 microscope equipped with the post-column Gatan BioQuantum energy filter connected to Gatan K2 Summit direct electron detector. Coma-free alignment was manually optimized and parallel illumination was examined prior to data collection. The data collection was semi-automatically executed by Leginon [40] version 3.1 and SerialEM [41] with the Gatan K2 Summit operating (Gatan Inc., CA, USA) in a super-resolution electron-counting mode and with the Gatan BioQuantum operating in the zero-loss imaging mode. Each movie contains 40 frames recorded with a total exposure time of 10 s with 250 ms per frame, with an accumulated dose of 44 electrons/Å^2. The calibrated physical pixel size and the super-resolution pixel size are 1.37 Å and 0.685 Å, respectively. The defocus was set in the range of -0.7 to -3.0 µm during data collection of a total of 44,664 movies.

Cryo-EM data processing

Cryo-EM data pre-processing, including motion correction, micrograph screening, and particle picking by deep-learning-based algorithm, was conducted as previously described [2]. Prior to 3D classification by the procedure of AlphaCryo4D, only 2D classification by ROME [42] and removal of apparent contaminants were conducted as previously reported [2]. All particles were binned twice down to 300 × 300 pixels prior to data processing. The general procedure of 3D classification by AlphaCryo4D has been described in detail elsewhere [12], which includes the

steps of *M*-fold particle shuffling and Bayesian bootstrapping of volume data, deep manifold learning of energy landscape, and energy-based particle voting. To proceed with AlphaCryo4D, the whole dataset was randomly split into 32 sub-datasets, each with 100,000 particles. 3-fold particle shuffling and volume bootstrapping procedures were conducted on each sub-dataset, which enhanced the total data size to 3,254,352 × 3 particles and yielded 32 × 4 × 10 = 1280 volumes in total that were used to compute the energy landscape of the substrate-engaged proteasome (Fig. 1B, C). Particle-voting on this energy landscape detected previously missing states (such as $E_{D0.2}$, $E_{D0.3}$, $E_{D1.1}$ and $E_{D1.2}$, etc.) and yielded previously determined states with improved map quality, such as E_{A2} and E_{D2}. Guided by this initial discovery of new conformational states, extensive focused 3D classifications in AlphaCryo4D were performed, as described in detail in the next section, to visualize novel ubiquitin-binding sites and hidden dynamics of proteasomal AAA-ATPase motor during single-nucleotide exchange. The Bayesian clustering during the bootstrap step and the final refinement of classified particles were conducted in RELION [10].

Focused 3D classification and reconstruction

To obtain a more complete picture of the conformational states of the substrate-engaged proteasome, we conducted focused 3D classification using AlphaCryo4D on several local regions on the overall energy landscape computed in the previous steps of analysis. To study the local features of ubiquitin binding and substrate interactions, 3D masks corresponding to the local densities of interest were applied to all bootstrapped volume data, and the subset of interest on the energy landscape was extracted to conduct a zoomed-in analysis. In this case, the energy landscape calculated via masked volumes is expected to reflect the local structural variations of interest (Fig. S2).

To detect novel ubiquitin-binding sites in the proteasome, local masks were applied to focus on the features around the putative locations of ubiquitin chains. After particle shuffling for M = 3 folds, 732,666 × 3 particle images of the state E_A and E_B were utilized for computing the zoomed-in energy landscape with 280 volumes. These particles were first aligned in a consensus alignment without any mask or resolution limit to restrict alignment parameter calculations. On the focused energy landscape, only the region within the soft mask of the ubiquitin and its binding site was applied in the volume bootstrapping step, with the signals of the CP of the proteasome subtracted from each raw particle when dealing with the state E_A and E_B.

Accordingly, the particle diameter applied in the CP subtracted proteasome was shrunk to 274 Å, instead of 411 Å in the complete 26S proteasome. When the region around RPN2 masked for focused 3D classification to search for potential ubiquitin-like features, the total particle numbers were enhanced to 208,700 × 3 (E_{A1}), 240,475 × 3 (E_{A2}), 260,021 × 3 (E_B) and 129,492 × 3 (E_{D2}). The energy landscapes were computed with 80, 80, 80 and 40 masked volumes bootstrapped by upper resolution bounds of 15, 15, 15 and 20 Å, respectively. Several 3D classes with the particle numbers of 192,219, 2,539, 3,982 and 609 were obtained by the particle-voting procedure on the zoomed-in energy landscape of state E_{A1}, while the case of state E_{A2} gave four classes containing 147,108, 5,842, 5,500 and 33,083 particles (Fig. S2, F and G). Likewise, the particle numbers of 3D classes on the zoomed-in energy landscape of state E_B were 173,931, 5,117 and 9,754 (Fig. S2H); and the zoomed-in energy landscape of state E_{D2} resulted in four classes having 61,580, 6,192, 24,674 and 1,673 particles, respectively (Fig. S2I).

To detect and improve RPN1-bound ubiquitin densities, a local mask around RPN1 was used, with 15 Å resolution limit applied in the volume-bootstrapping step. The zoomed-in energy landscape around state E_{A1} computed from 120 volumes resulted in five classes with the particle numbers of 128,161, 7,754, 6,557, 6,826 and 3,937. Five 3D classes on the zoomed-in energy landscape of E_{A2}, also computed with 120 volumes, including 81,612, 27,483, 12,247, 26,029 and 12,777 particles, were analyzed (Fig. S2C, D). Based on the subclass of E_{A1} that contains 128,161 particles, we continued to refine the alignment locally to the component of RPN1. Then a 3D mask around RPN1 was applied to bootstrap 40 volumes with 20 Å resolution limit in the expectation step. After voting on the energy landscape, two classes with the particles number of 43,957 and 43,620 were generated, one of which exhibited a ubiquitin-like density on the T1 site of RPN1 and was designated state $E_{A1.1}$ (Fig. S2E). In the case of the ubiquitin-binding site on the α5 subunit of the CP, five 3D classes with the particle numbers of 136,071, 19,351, 4,100, 131,570 and 4,932 were generated by particle voting on the zoomed-in energy landscape of E_D computed with 120 volumes, which were bootstrapped from 326,409 × 3 particles via the mask containing this ubiquitin-binding site and no high-resolution restriction (Fig. S2J). During these analyses, all masked volume-bootstrapping steps were performed for 50 iterations, whereas unmasked 3D classifications were performed for 30 iterations with the initial reference low-pass filtered to 60 Å. At the last step, 3D refinement and reconstruction of each class was done to finalize the high-resolution structure determination.

To exploit intermediate conformations during the substrate translocation initiation, a zoomed-

in energy landscape was computed for the focused 3D classification procedure. First, the energy landscape of all 3,254,352 × 3 particles was computed and analyzed. Then the zoomed-in energy landscape of the states E_C and E_D containing 2,521,686 × 3 particle images were computed from 1,000 volumes that were bootstrapped without masking or resolution restriction (Fig. S2B). For further zoomed-in processes, the particles around the translocation initiation states (from E_C to E_{D0}) on this energy landscape were extracted by the distance-based 3D classification strategy. Subsequently, these particles were shuffled 3 times to the number of 410,098 × 3 to generate 160 volumes for the computation of the zoomed-in energy landscape with the RP mask and no resolution restriction imposed in the volume bootstrapping, which led to the 3D classes E_{C1}, E_{C2}, and $E_{D0.1}$ with the particle number of 68,506, 77,545, and 25,755, respectively (Fig. S2K). Moreover, state $E_{D0.3}$ came from two clusters with 86,415 and 83,107 particles merged together on the unmasked energy landscape, and the classes $E_{D0.2}$ and $E_{D1.1}$ including 105,081 and 51,808 particles, respectively, were also obtained after particle voting on this energy landscape. In the local areas of the state E_{D1} on the energy landscape, two clusters with 80,841 and 94,273 particles were merged to reconstruct the high-resolution structure of state $E_{D1.2}$, while states $E_{D1.1}$ and $E_{D1.3}$ with 40,696 and 145,506 particles, respectively, were separated out from the clusters on this unmasked energy landscape. In these studies, the maximum number of iterations in the 3D classifications of volume bootstrapping was 30, with the initial reference low-pass filtered to 60 Å routinely.

Beyond resolving significantly more conformers from the same dataset, AlphaCryo4D also allows us to push the envelope of the achievable resolution of dynamic components and key metastable states due to its advantage in keeping more particles without sacrificing conformational homogeneity. To improve the resolution of state E_{D2}, a rigorous particle voting procedure was performed in each cluster on the unmasked energy landscape (Fig. 2B, C), which improved the overall quality of particle dataset while classifying significantly more particles to this E_{D2} conformational class. Next, we conducted a focused 3D classification by using AlphaCryo4D to calculate the zoomed-in energy landscape of the entire E_{D2} dataset to detect any potential conformational changes within this class. Five clusters near the energy wells of state E_{D2}, whose particle numbers were 154,661, 243,799, 198,221, 179,257 and 106,047, were investigated by separate high-resolution 3D refinement and reconstruction. The resulting refined density maps at 3.0-3.5 Å were found to be all in an identical conformation as previously reported [2]. These five clusters were then merged together to refine a final 3D density map of state

E_{D2}. Thus, AlphaCryo4D allowed us to obtain a new E_{D2} dataset that appears to retain high conformational homogeneity while including three-fold more particles, which improves its 3D reconstruction from the previously published resolution at 3.2 Å to 2.5 Å, measured by gold-standard Fourier shell correlation (FSC) at 0.143-cutoff. By contrast, in our previous work [2], we had attempted to include more E_{D2}-compatible 3D classes from 3D maximum-likelihood classification in RELION, which resulted in lower resolution at 3.3 Å by doubling the E_{D2} dataset, indicating that including more particles by conventional methods gave rise to higher heterogeneity in the class presumably due to increased misclassification.

Analysis of substrate-free proteasome dataset

The dataset of the ATPγS-bound human 26S proteasome [8] contains 455,680 particles and was previously used to determine six coexisting conformational states of the substrate-free proteasome. The cryo-EM data were recorded with the super-resolution pixel size of 0.75 Å. The dataset is publicly available in the Electron Microscopy Public Image Archive (EMPIAR) database at EMDataResource.org (accession code EMPIAR-10090). The size of ATPγS-bound proteasome dataset was enhanced to 455,680 × 3 particles and was divided into 4 sub-datasets for volume bootstrapping. 160 volumes were bootstrapped to compute the energy landscape the substrate-free ATPγS-bound proteasome (Fig. 1A).

Analysis of doubly and singly capped proteasomes

Previous cryo-EM analyses of the proteasomal states were nearly all focused on the RP-CP subcomplex, by converting double-cap (DC) proteasome particles to pseudo single-cap (SC) proteasome, which was always mixed with true SC particle images, in order to improve the resolution with conventional methods [2,7-9,43,44]. As a result, the cryo-EM density of the CP gate in the other side distal to the RP in the RP-CP subcomplex has been an average of closed and open states with unknown stoichiometry. This has made it impossible to appreciate the difference between the SC and DC proteasomes in the previous studies. To remove this ambiguity and local conformational heterogeneity, we reconstructed 8 density maps of the pure SC proteasome and 36 density maps of the pure DC proteasome including all possible combination of the 8 major RP-CP states by AlphaCryo4D classification. Both CP gates in the 44 states of the pure SC and DC reconstructions are expected to be in a homogeneous conformation. To reconstruct the density maps of the DC proteasome, pseudo-SC particles corresponding to 8 different RP states

were voted out from the energy landscape, with the sub-states belonging to the same RP state combined together. Then the particles with one RP-CP were extracted from the particle stacks to refine the RP and CP density maps with RP and CP contour masks applied in the local search step in RELION, respectively. For each reconstruction of the DC proteasome, two RP density maps on the opposite sides of the CP were expanded from 600 × 600 × 600 pixels to 800 × 800 × 800 pixels using RELION and aligned by the density of CP in Chimera before being merged together. All 8 SC and 29 DC states were refined to 3.7-4.7 Å, with 7 DC states limited to 6-9 Å due to their transient nature and extremely low population (Fig. S4, B, E and F).

Counting the particle number of the 36 states of the DC proteasome, we could derive the distribution of the DC states, p_{ij}, and that of RP p_i only, where i and j denote the states of two RPs in the same DC proteasome. Based on the value of p_{ij} and total particle number, the experimental state distribution of the DC proteasome was plotted for quantitative analysis in two dimensions with respect to the states of two RPs in the same DC proteasome (Fig. 1E). If the states of two RPs of the same DC proteasome were independent of each other, the predicted state distribution $\tilde{p}_{ij}$ of the DC proteasome can be calculated as:

$$\tilde{p}_{ij} = p_i \cdot p_j$$

By using the total particle number and experimentally measured p_i, the 2D state distribution of the DC proteasome with two uncoupled RPs was calculated for further comparison to investigate the conformational entanglement effect of the two RP states (Fig. 1, F-H).

Atomic model building and refinement

To build the initial atomic model of the newly discovered states, we use previously published proteasome structures as a starting model and then manually improved in Coot [45]. For the conformational states at resolutions lower than 5 Å, pseudo-atomic modelling was conducted in the following steps. First, each subunit was fitted as a rigid body, often in UCSF Chimera [46] as well as in Coot [45]. Then, the local structural domains and secondary structure elements were fitted as a rigid body, with linking loops flexibly fitted and modelled. After the mainchain structures were well fitted, no further fitting of the sidechain rotamers were pursued due to insufficient resolution except for energy minimization to remove unrealistic side-chain clashes. For the conformational states at resolutions higher than 5 Å, the mainchain traces were first fit in Coot. Then, extensive fitting, adjustment and optimization of sidechain rotamers were conducted

through local real-space refinement and manual rectification in Coot to the degree commensurate to the map resolution and quality. Atomic model refinement was conducted in Phenix [47] with its real-space refinement program [48]. We used both simulated annealing and global energy minimization with secondary structure, rotamer and Ramachandran constraints. Partial rebuilding, model correction and density-fitting improvement in Coot [45] were iterated after each round of atomic model refinement in Phenix [47]. The improved atomic models were then refined again in Phenix, followed by rebuilding in Coot [45]. The refinement and rebuilding cycle were often repeated for three rounds or until the model quality reached expectation (Table S1). All figures of structures were plotted in Chimera [46], PyMOL [49], or ChimeraX [50]. Local resolutions of cryo-EM density maps were evaluated using ResMap [51] or Bsoft Blocres program [52]. Structural alignment and comparison were performed in both PyMOL and Chimera. Electrostatic surfaces were calculated by APBS plugin [53] in PyMOL.

Given a large range of variations of local resolution in an expanded number of proteasome states and substates, as well as reconstructions with different stoichiometric ratios of RP versus CP, additional caveats and cautions were practiced in order to avoid misinterpretation and overall-fitting of atomic and pseudo-atomic models. First, when fitting the ubiquitin structure to the low-resolution local density, we consider both the existing homologous structural models as a modelling reference. Specifically, the fitting of diubiquitin to the density on RPN2 in states $E_{A2.1}$, $E_{A2.2}$, and $E_{B.2}$, we took the NMR structure of the yeast diubiquitin-bound Rpn1 T1 site as a reference, because of the high structural homology between RPN1 and RPN2. For the fitting of RPN10-bound density, we also took into account the electrostatic complementarity (Fig. S7H-O). Second, even within an overall high-resolution cryo-EM map, it often presents certain local densities at a lower local resolution due to poor occupancy of ligands or flexibility of less well-folded segments. For example, for the nucleotide in RPT5 in states $E_{D0.1}$ to $E_{D1.3}$, the high-resolution atomic model of ADP refined from other states like E_{C1} were used to fit these lower-local-resolution densities as a rigid body. However, the B-factors of the fitting atoms of ADP in RPT5 were reported to be more than twice higher than those of other nucleotides, indicating its partial occupancy or unstable binding. Thus, ADP fitting in RPT5 should be regarded tentative for the most or for the purpose of evaluation of nucleotide states rather than a reliable high-resolution atomic modeling. Similarly, the atomic models of these locally low-resolution features were only further adjusted with strict stereochemical constraints when the map resolution and features permit. Third, for the CP gate at the other side opposite to the RP in the RP-CP

subcomplex reconstructed from a mixture of true and pseudo-SC proteasomes (converted from the DC proteasomes), their local densities in states $E_{D0.1}$ to E_{D2} appear to be an average of the open and closed states, and thus are weaker in amplitude than the rest of the CP. In the atomic model building of these reconstructions, we chose to model them with the closed CP gate state as long as their density resolution allows atomic modeling. This issue was completely solved when we separated the 3D classes of the SC and DC proteasomes and reconstructed their states respectively, where the atomic modelling of any CP gate no longer suffers from such ambiguities.

Quantum mechanical calculation

First-principles calculations on the electronic structures of the substrate-proteolytic site complex were conducted as described. In brevity, local coordinates of the substrate-bound CP near the catalytic site at residue Thr1 were taken out from the experimentally fitted atomic models of CP and the substrate, in both E_A and E_{D2} conformations. The proteolytic site is the Thr1 residue at the β2 subunit but extended along the chain for another two residues to make the system less finite. Hydrogen atoms are added to complete the residues, and to saturate the system boundaries. The structural relaxations and electronic structure calculation were carried out using the density functional theory (DFT) method with norm-conserving pseudopotentials as implemented in the SIESTA code [54]. To set up the initial conditions of the finite system, the hydrogen atom in the hydroxy group of Thr1 is manually relocated to the nearby nitrogen atom in the same residue, corresponding to the initial proton transfer in the catalytic mechanism. All hydrogen atoms were subsequently optimized without symmetric restrictions using the conjugate gradient algorithm and a 0.04 eV/Å maximum force convergence criterion while keeping the rest of the system fixed. The generalized gradient approximation (GGA) exchange-correlation density functional PBE [29] was employed together with a double-zeta plus polarization basis set, and a mesh cutoff of 200 Ry (corresponds to 0.23 Å smallest grid size). The charge density and charge density difference contour maps were plotted with the Siesta Utility programs denchar and Python, respectively.

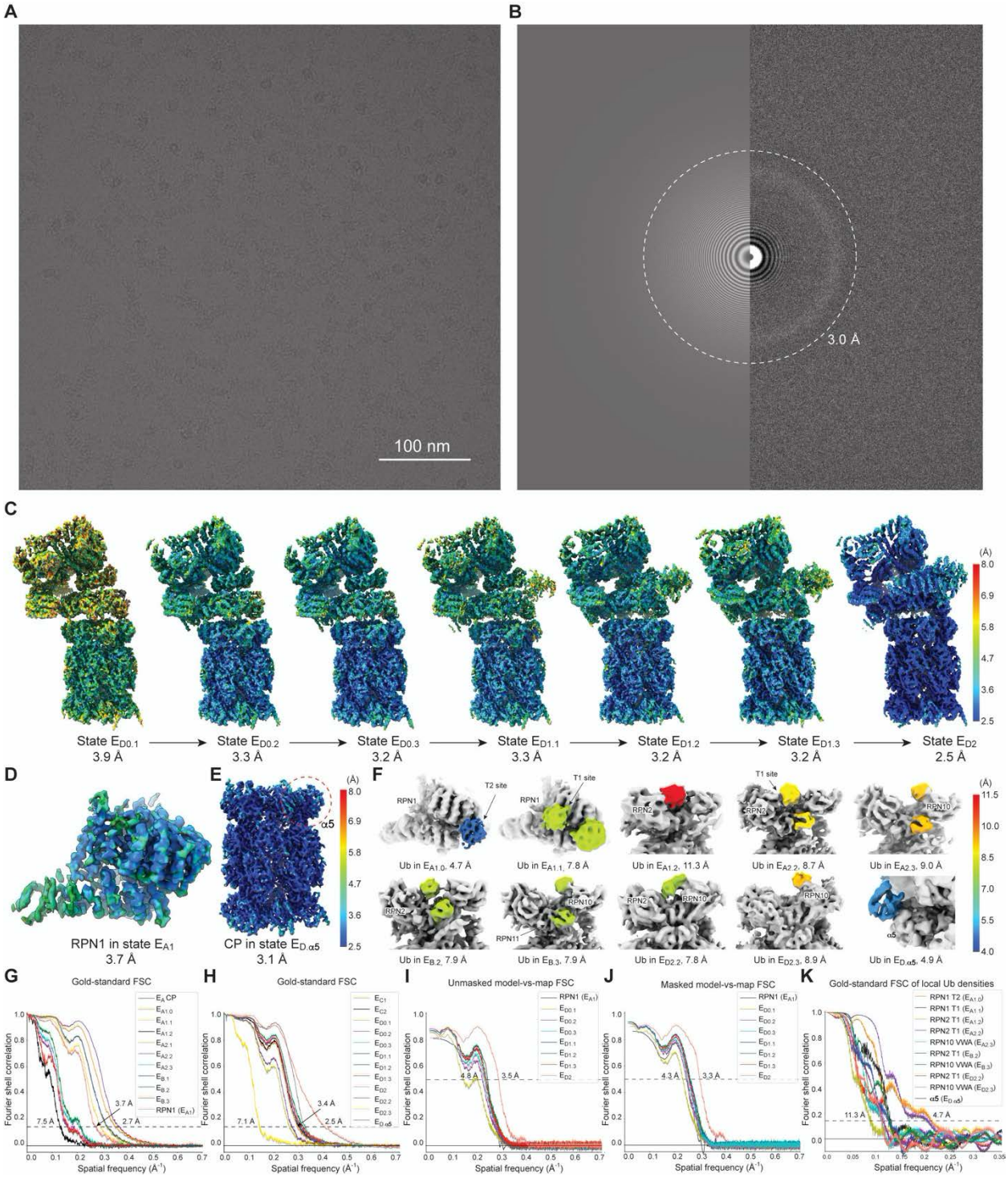

Fig. S1. Cryo-EM reconstructions of novel proteasomal states.

(**A**) A typical raw micrograph of the substrate-bound human proteasome recorded under the super-resolution counting mode of Gatan K2 Summit direct electron detector. (**B**) The power spectrum of the micrograph shown in panel (**A**), with the 3 Å resolution ring marked by the dashed circle. (**C**) Local resolution measurement of the seven sequential intermediate states during translocation initiation measured by ResMap [51]. (**D**) The local resolution measurement of the improved density map of RPN1 in state E_{A1} by ResMap. (**E**) The local resolution

measurement of the CP in state $E_{D.\alpha5}$ by ResMap. (**F**) Putative ubiquitin (Ub) densities in ten distinct states colored by the local gold-standard FSC resolutions of the Ub densities measured in panel (**K**), with the rest of densities shown in light grey. The color bar for local resolutions is shown on the right in panels in **C**-**F**, where panels (**D**) and (**E**) share the same resolution color bar. (**G** and **H**) The gold-standard FSC plots of the 20 conformers of the RP-CP subcomplex. (**I** and **J**) The FSC curves calculated between the experimental cryo-EM maps and their corresponding atomic models for newly discovered states or improved states without (**I**) and with (**J**) masking in the calculations. **(K**) The gold-standard FSC curves calculated between the two half maps with local Ub masks in ten distinct states.

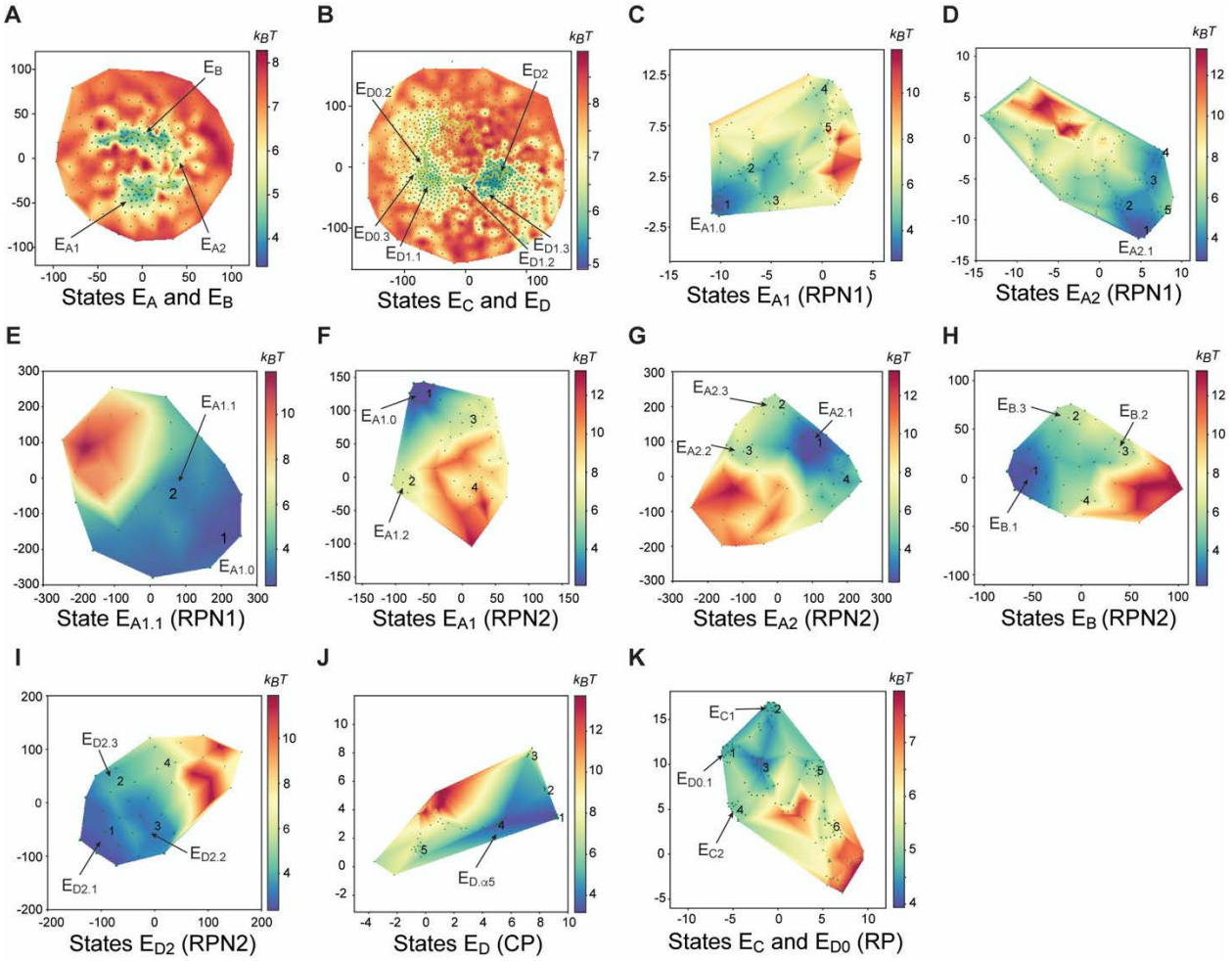

Fig. S2. Energy landscapes of substrate-engaged human proteasome computed to guide in-depth focused 3D classification.

(**A**-**K**) The zoomed-in energy landscape used for focused 3D classifications using program AlphaCryo4D for finding new states of the substrate-bound proteasome. The zoomed-in local energy landscapes of the substrate-bound human proteasome by AlphaCryo4D are shown, respectively, for state E_A and E_B in (**A**), for states E_C and E_D in (**B**), for states E_{A1} and E_{A2} with 3D mask focusing on RPN1 in (**C**) and (**D**) to visualize the RPN1 T2 site, in (**E**) to visualize the RPN1 T1 site, for states E_{A1}, E_{A2}, E_B and E_{D2} with 3D masking around RPN2 in (**F**), (**G**), (**H**) and (**I**) to visualize the RPN2 T1 site, for combined states of E_D with masking around CP (**J**) to visualize the ubiquitin-binding sites on the α5-subunit, and for combined states of E_C and E_{D0} with 3D mask focusing on the RP (**K**) to visualize the ATPase motor intermediates during translocation initiation.

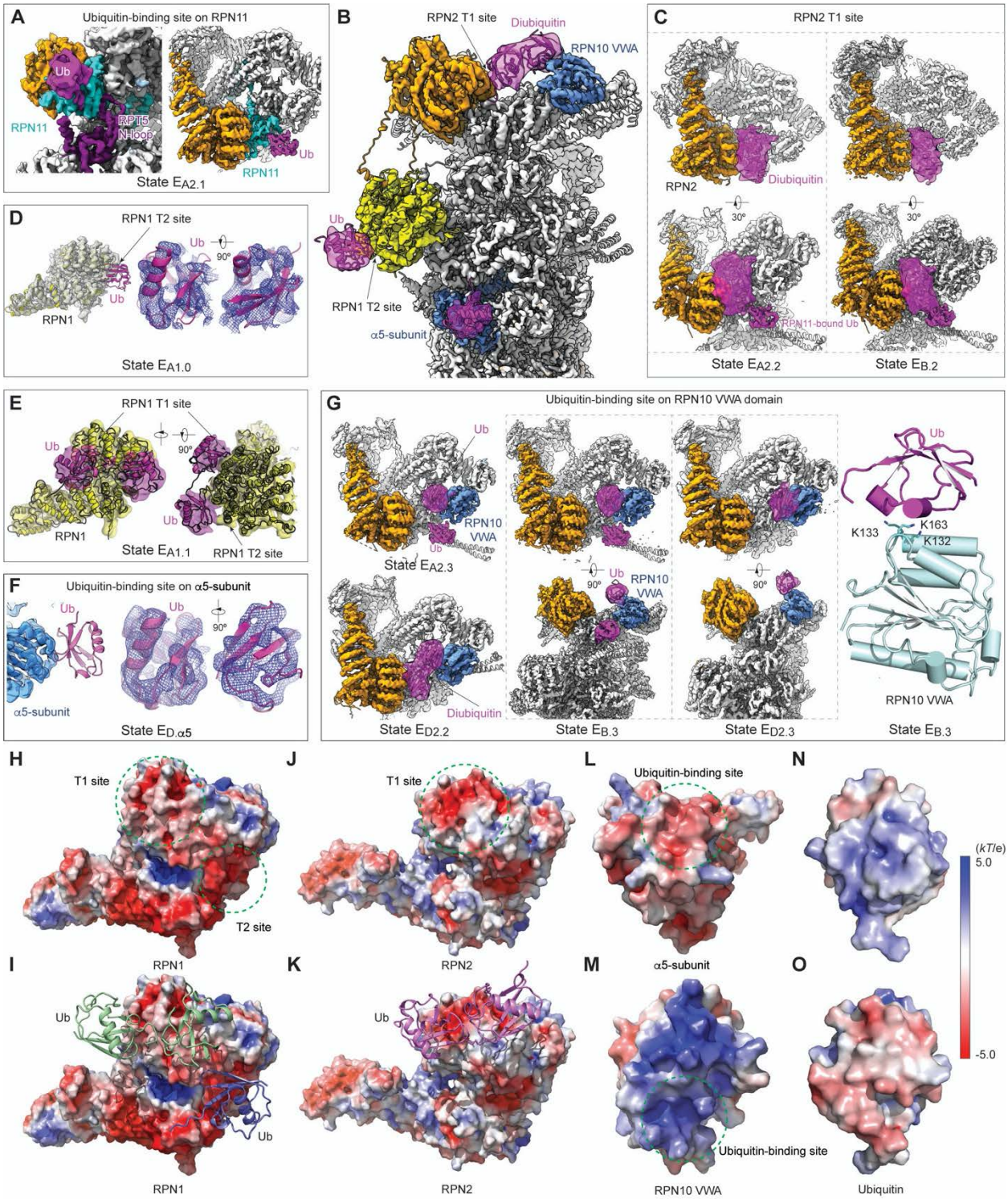

Fig. S3. Structural analysis of ubiquitin-proteasome interactions.
(**A**) Cryo-EM density of state $E_{A2.1}$ shows the improved resolution localized around RPN11-bound ubiquitin and the RPT5 N-loop as compared to previously determined state E_{A2}. Ub, ubiquitin. (**B**) High-resolution cryo-EM density of state E_{D2} superimposed with Ub densities in state $E_{D2.2}$ highlighted as magenta, showing that a diubiquitin chain in contact with both the RPN2 T1 site and the RPN10 VWA domain. The atomic model of state $E_{D2.2}$ in cartoon representation is superimposed with the density map. (**C**) Closeup view of the RPN2-bound

diubiquitin density in states $E_{A2.2}$ (left column) and $E_{B.2}$ (right column) in two different perspectives, with the lower row showing the coexistence of the RPN11-bound Ub and the RPN2-bound diubiquitin. (**D**) The refined 3.7-Å RPN1 density map from state $E_{A1.0}$ in transparent surface representation superimposed with its atomic model with Ub bound near the RPN1 T2 site. The right insets show the local cryo-EM density of the RPN1-bound Ub in state $E_{A1.0}$ superimposed with its atomic model in cartoon representation. (**E**) Cryo-EM density of state $E_{A1.1}$ in transparent surface representation superimposed with its pseudo-atomic model of cartoon representation, with one Ub moiety bound to the RPN1 T1 site and another Ub near the T2 site. The two Ub moieties appear to be compatible with a Lys63-linked diubiquitin model in an extended conformation, which is supported by the low-level density connecting the two Ub densities. (**F**) Closeup view of the α5-subunit bound with Ub in the refined 3.1-Å density map of state $E_{D.\alpha5}$. The cryo-EM density of the α5-subunit is shown in transparent surface representation superimposed with its atomic model. The right insets show the local cryo-EM density of the α5-bound Ub of state $E_{D.\alpha5}$ in blue mesh representation low-pass filtered at 4.9 Å and superimposed with its atomic model in cartoon representation. (**G**) Closeup views of the RPN10-bound Ub densities in states $E_{A2.3}$, $E_{B.3}$, and $E_{D2.3}$. Right panel shows the pseudo-atomic model of Ub interaction with the RPN10 VWA domain. The residues of RPN10 in contact with ubiquitin are labeled and shown as stick representation. Note that the Ub-binding mode of the RPN10 VWA domain here is different from that previously observed in the crystal structure of ubiquitylated yeast Rpn10 that also suggested Rpn10 VWA domain as a ubiquitin-binding domain[15]. In the yeast Rpn10, Ub binds to the Rpn10 surface area interfacing Rpn9 and sterically occluded Rpn10 incorporation into the proteasome [15]. (**H-M**) The electrostatic surfaces of the RPN1, RPN2, RPN10 and α5 subunits show the charge complementarity of the RPN1 T1 and T2 (panels **H**, **I**), RPN2 T1 (panels **J**, **K**) and α5-binding (panel **L**) sites of Ub interactions are all acidic. By contrast, the RPN10 VWA site is basic (panel **M**), suggesting that it may bind the acidic side of Ub. (**N** and **O)**, The electrostatic surface of Ub shows one side that is basic (panel **N**) and the other side acidic (panel **O**).

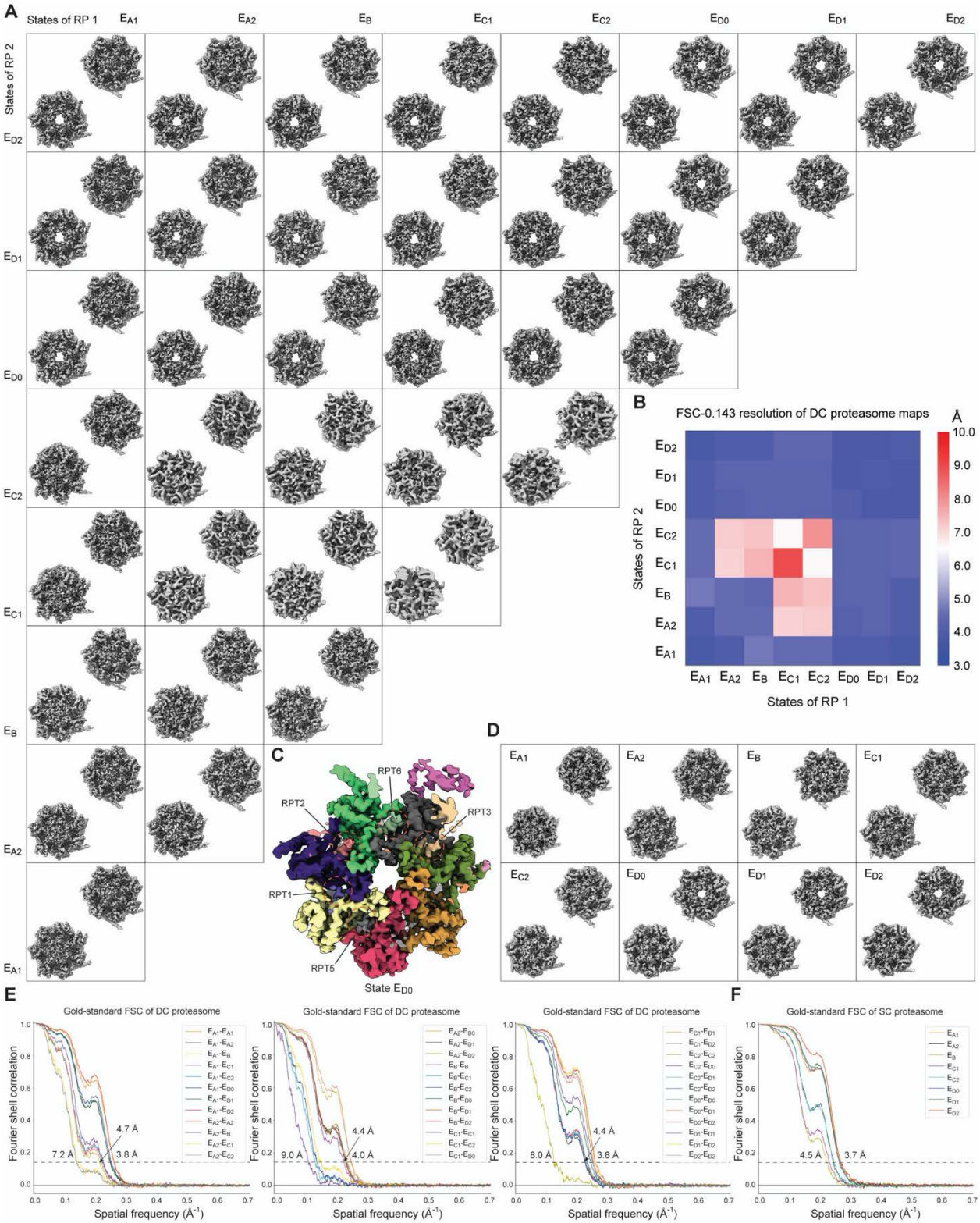

Fig. S4. Cryo-EM analysis of the doubly capped (DC) and singly capped (SC) proteasomes. (**A**) The CP gate states of the 36 density maps of distinct conformational states of the DC proteasome. All density maps shown are low-pass filtered to their measured resolutions by the gold-standard FSC-0.143 cutoff without amplitude correction of B-factors. (**B**) The FSC-0.143 resolution matrix of the 36 DC proteasomal states. (**C**) The RP-CP interface of state E_{D0},

showing five C-terminal tails inserted into the inter-subunit surface pockets of α-ring. (**D**) The CP gate states of the 8 density maps of distinct conformational states of the SC proteasome, with the upper right image in each panel corresponding to the RP-proximal side of the CP gate, and the lower left image to the RP-distal side of the CP gate. (**E**) The gold-standard FSC plots of the 36 conformers of the DC proteasome. (**F**) The gold-standard FSC plots of the 8 conformers of the SC proteasome.

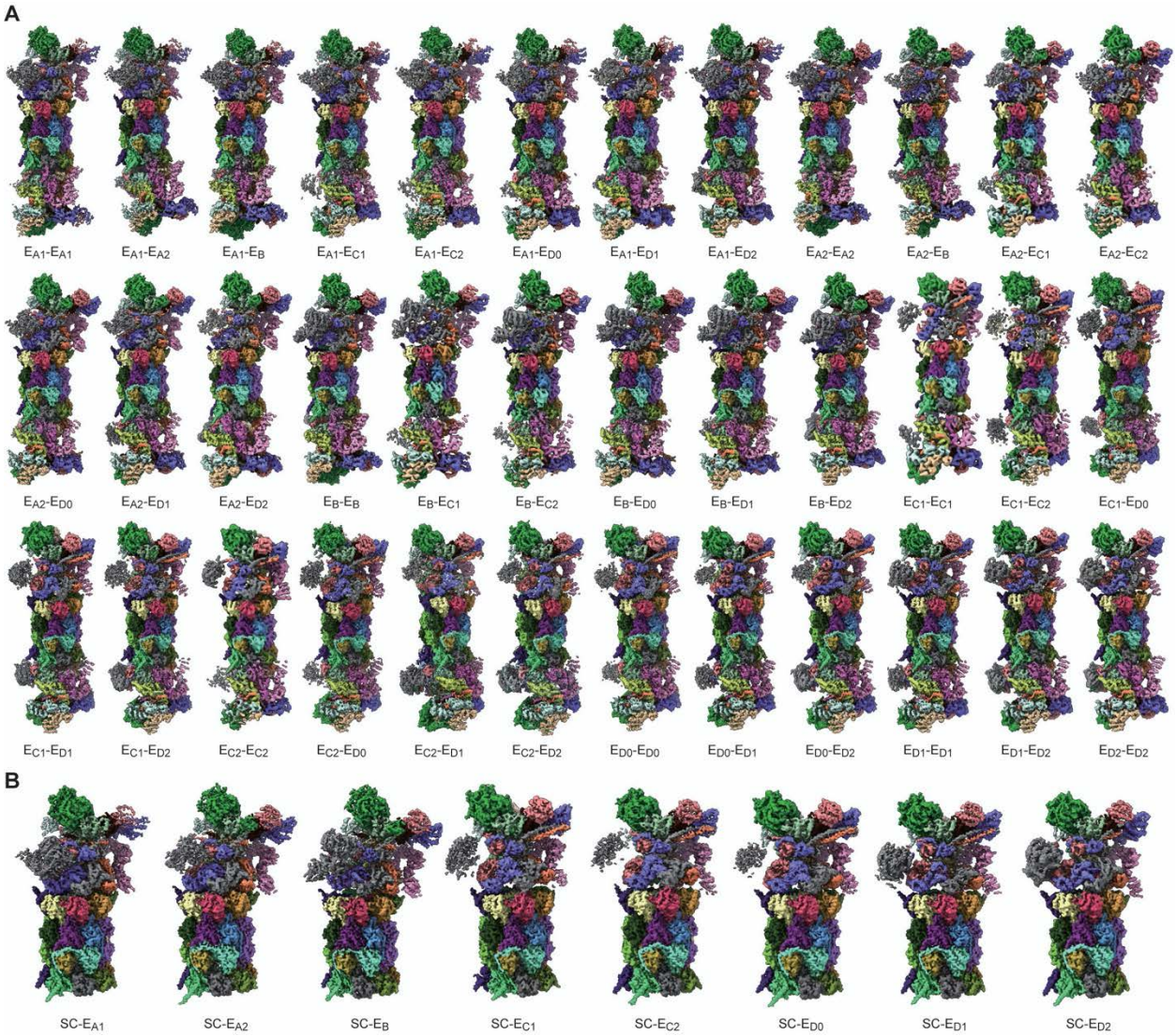

Fig. S5. Cryo-EM reconstructions of the SC and DC proteasomes.
(**A**) The density maps of 36 distinct conformers of the DC proteasome classified by AlphaCryo4D. (**B**) The density maps of 8 distinct conformers of the SC proteasome. All maps are filtered to their respective gold-standard FSC-0.143 resolutions without amplitude correction of B-factors and are differentially colored by their subunits in ChimeraX [50].

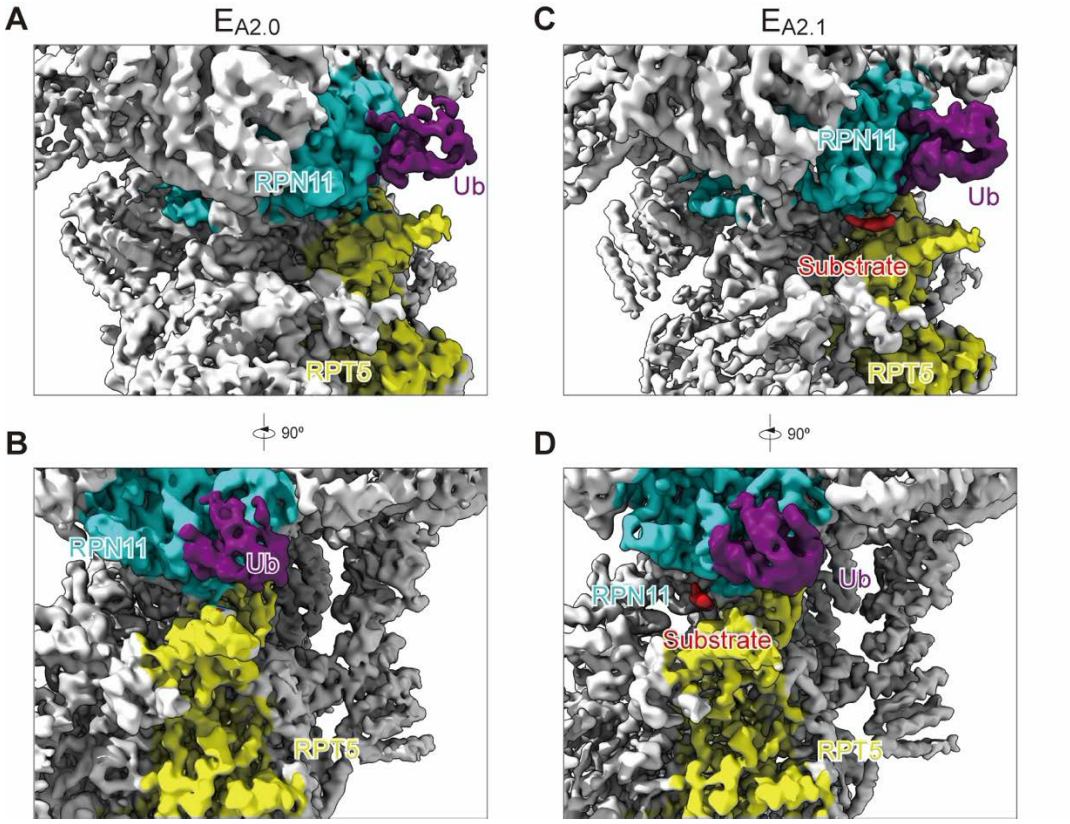

Fig. S6. **A newly identified intermediate state $E_{A2.1}$ between state E_{A2} and E_B.** (**A** and **B**) Close-up view of the ubiquitin interaction with RPN11 in state $E_{A2.0}$ shown in two orthogonal perspectives. The RP conformation in state $E_{A2.0}$ is identical to the previously reported state E_{A2}, but the local resolution of the RPN11-bound ubiquitin is improved in state $E_{A2.0}$ compared to that of state E_{A2}. (**C** and **D**) Close-up view of the ubiquitin interaction with RPN11 in state $E_{A2.1}$ shown in two orthogonal perspectives, where a short segment of substrate density (highlighted as red) is observed to bind the cleft between RPN11 and RPT4 near the active site of RPN11. While both states $E_{A2.0}$ and $E_{A2.1}$ show very similar AAA-ATPase conformations that are not yet engaged with the substrate, the lid is also rotated by a small angle in state $E_{A2.1}$ relative to state $E_{A2.0}$. The RPT5 N-loop, C-terminal strand of ubiquitin and insert-1 hairpin of RPN11 already form a four-stranded β-sheet in both states $E_{A2.0}$ and $E_{A2.1}$, suggesting that the formation of the four-stranded β-sheet on RPN11-ubiquitin interface proceeds the substrate binding to the RPN11-RPT4 cleft. These features indicate that state $E_{A2.1}$ represent an intermediate between states E_{A2} and E_B that proceeds the substrate insertion into the AAA-ATPase and deubiquitylation.

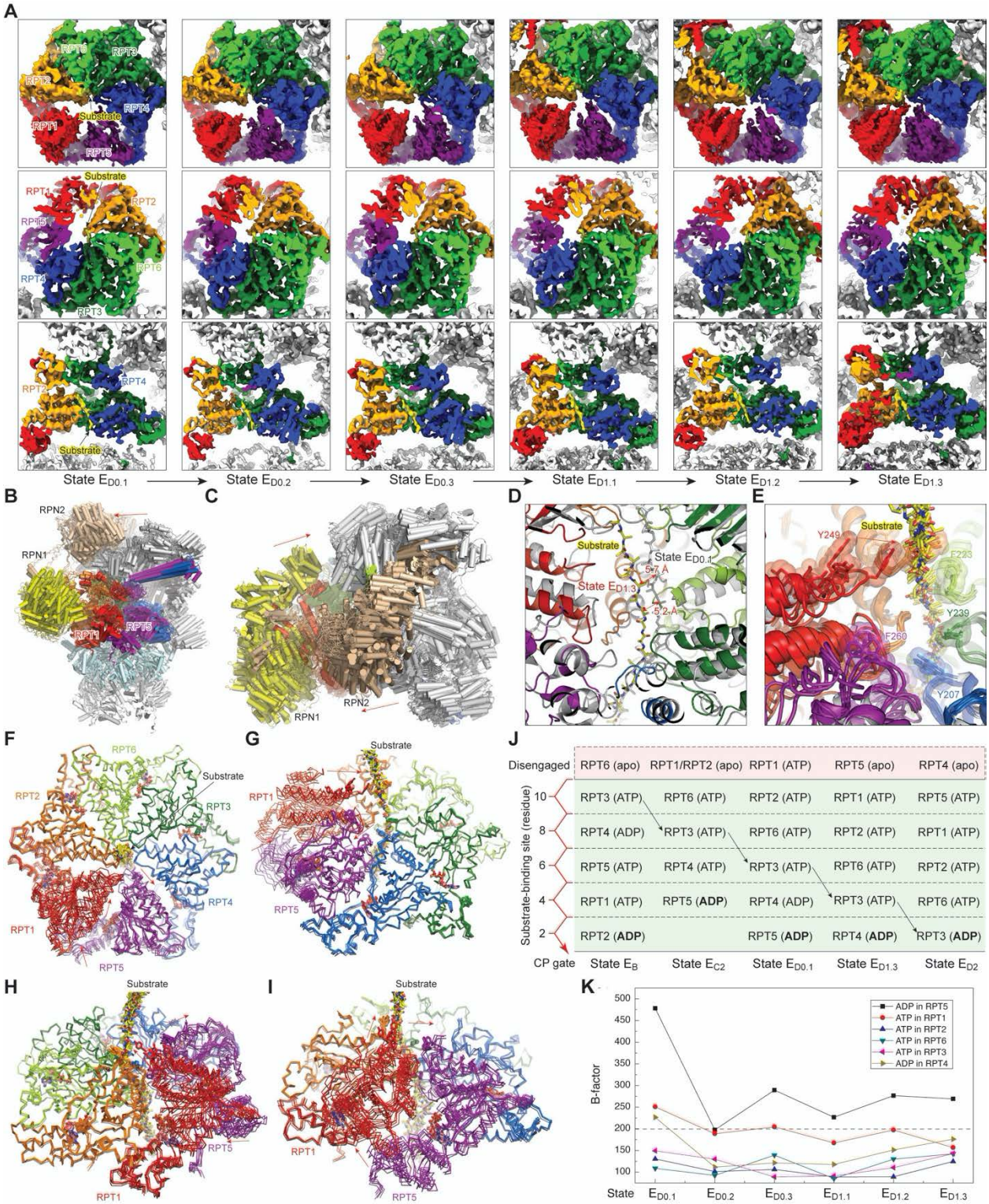

Fig. S7. Cryo-EM structures of six sequential intermediate states between E_{C2} and $E_{D1.3}$. (**A**) Cryo-EM density maps of the AAA-ATPase motor in six sequential states from $E_{D0.1}$ to $E_{D1.3}$ from three different viewing angles. The substrate densities are highlighted in yellow. (**B**) Superposition of the seven structures in cartoon representations from a lateral perspective, with all structures aligned together based on their CP structures, showing the RP movements relative to the CP. (**C**) Superposition of the six structures in cartoon representations from a top-view

perspective, with all structures aligned together based on their CP structures. (**D**) Measurement of substrate movement between states $E_{D0.1}$ and $E_{D1.3}$ relative to the CP, suggesting that the substrate is translated ~5-6 Å during the process of ADP release in RPT5 and of RPT1 re-association with the substrate. (**E**) The closeup view of the pore-1 loop interaction with the substrate with all six states superimposed after aligned together based on the structures of RPT3, RPT4 and RPT6. (**F**-**I**) Superposition of the six sequential states with their RPT3, RPT4 and RPT6 aligned together. Red arrows show the direction of subunit movements in RPT1 and RPT5. (**J**) A diagram illustrating the axial stepping of the substrate-interacting pore-1 loops that is coupled with ATP hydrolysis in the RPT subunits, revised from the previously published results based on the present study [2]. (**K**) The average B-factor of the nucleotides in six ATPase subunits fitted to the cryo-EM densities of states $E_{D0.1}$ to $E_{D1.3}$ and computed by Phenix in real-space refinement procedure [47]. The high B-factor of ADP fitted in RPT5 indicates its low occupancy or unstable association and suggests that the ADP in RPT5 undergoes dissociation in this process.

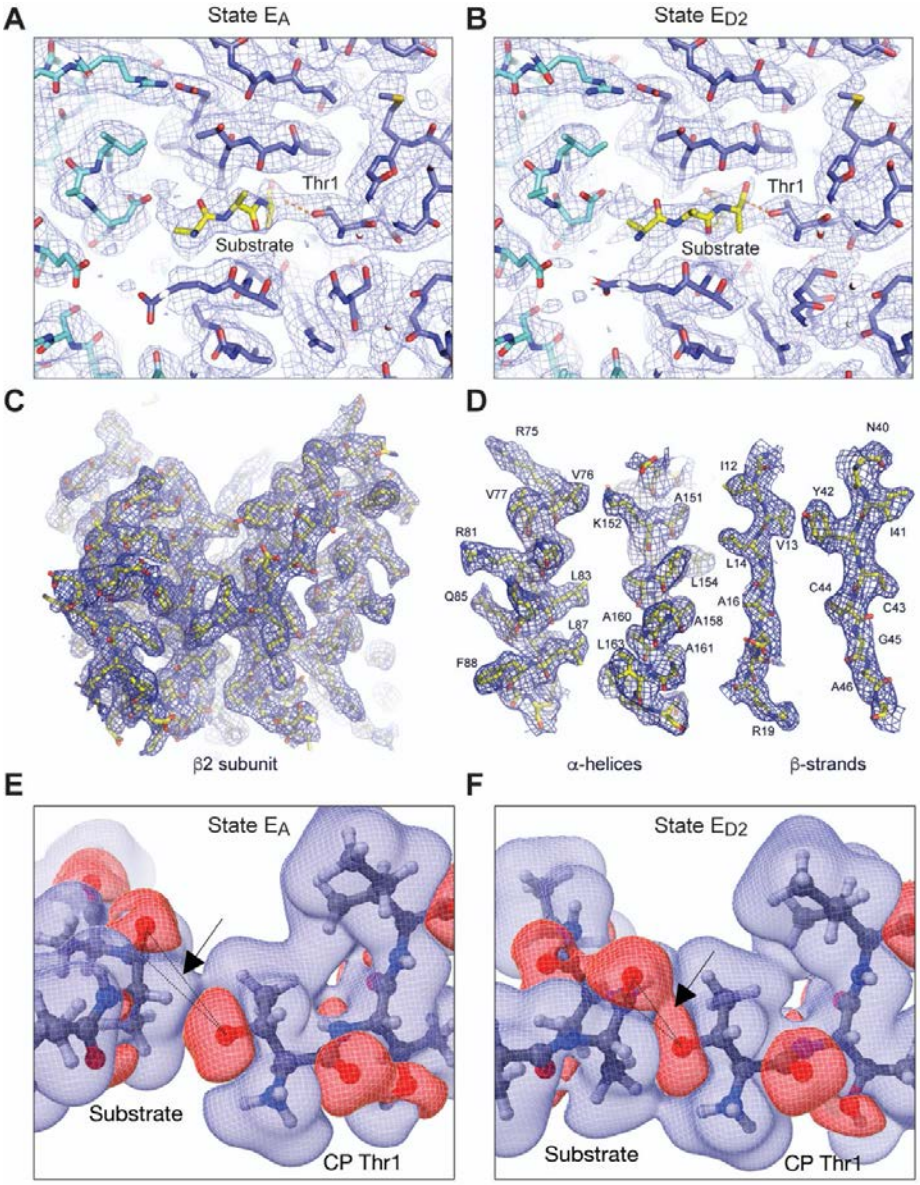

Fig. S8. Substrate interactions with the proteolytic sites in two major CP states. (**A**) The atomic model of substrate-bound catalytic active site of the β2-subunit superimposed with the cryo-EM density map of state E_A at 2.7 Å. (**B**) The atomic model of substrate-bound catalytic active site of the β2-subunit superimposed with the cryo-EM density map of state E_{D2} at 2.5 Å. (**C**) Closeup view of the cryo-EM density of the β2-subunit in state E_{D2} shown in blue mesh representation superimposed with the atomic model. (**D**) Cryo-EM densities of the typical secondary structure elements in the β2-subunit of state E_{D2} shown as blue mesh representation superimposed with their corresponding atomic models. The residues are labelled. The densities are shown at 4s level and exhibit structural features consistent with 2.5-Å resolution. (**E** and **F**) 3D Charge density difference map showing the interactions between the substrate polypeptide and the residue Thr1 in the β2-subunit of the CP. The red color labels an iso-surface of charge increase while the blue color labels charge decrease, both at a level of 0.04 e/Å^2. The charge increase region at the Thr1-Oγ atom is elongated sideways, and the substrate position in the E_{D2} state provides both a proximity and orientation alignment for the pair to interact. The dashed line

triangle shows the plane in which the detailed charge difference contours are plotted as shown in Fig. 5C and D.

Table S1. Cryo-EM data collection, refinement and validation statistics.

	$E_{D0.1}$	$E_{D0.2}$	$E_{D0.3}$	$E_{D1.1}$	$E_{D1.2}$	$E_{D1.3}$	$E_{D2.1}$
Data collection and processing							
Magnification	105,000	105,000	105,000	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300	300	300	300
Electron exposure (e–/Å^2)	44	44	44	44	44	44	44
Defocus range (μm)	-0.6 to -3.5	-0.6 to -3.5	-0.6 to -3.5	-0.6 to -3.5	-0.6 to -3.5	-0.6 to -3.5	-0.6 to -3.5
Pixel size (Å)	0.685	0.685	0.685	0.685	0.685	0.685	0.685
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1
Initial particle images (no.)	3,254,352	3,254,352	3,254,352	3,254,352	3,254,352	3,254,352	3,254,352
Final particle images (no.)	25,755	105,081	169,522	92,777	175,114	146,506	856,683
Map resolution (Å)	3.9	3.3	3.2	3.3	3.2	3.2	2.5
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.0-8.0	2.5-8.0	2.5-8.0	2.5-8.0	2.5-8.0	2.5-8.0	2.5-4.8
Refinement							
Initial model used			6MSH			6MSJ	6MSK
Model resolution (Å)	4.3	4.1	4.0	3.9	3.8	3.9	3.2
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Model resolution range (Å)	3.0-8.0	3.0-8.0	3.0-8.0	3.0-8.0	3.0-8.0	3.0-8.0	2.5-5.8
Map sharpening *B* factor (Å^2)	-20	-20	-20	-20	-30	-30	-20
Model composition							
Non-hydrogen atoms	106475	106475	106475	106475	106448	106448	106499
Protein residues	13540	13540	13540	13540	13540	13540	13540
Ligands	11	11	11	11	10	10	11
B factors (Å^2)							
Protein	155.88	155.03	155.53	126.38	112.34	133.02	88.54
Ligand	268.45	173.9	201.79	134.06	115.01	144.41	88.86
R.m.s. deviations							
Bond lengths (Å)	0.006	0.008	0.007	0.006	0.007	0.007	0.007
Bond angles (°)	1.022	1.095	1.076	1.041	1.047	1.072	1.074
Validation							
MolProbity score	1.77	1.82	1.77	1.82	1.45	1.78	1.75
Clashscore	5.05	5.58	5.08	5.57	5.13	5.14	4.49
Poor rotamers (%)	0.29	0.47	0.39	0.44	0.35	0.36	0.37
Ramachandran plot							
Favored (%)	91.56	90.99	91.47	90.98	91.71	91.5	91.24
Allowed (%)	8.15	8.73	8.24	8.75	7.91	8.23	8.47
Disallowed (%)	0.29	0.28	0.28	0.27	0.39	0.28	0.29