Supplementary Materials for

2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition

Soojay Banerjee, Alberto Bartesaghi, Alan Merk, Prashant Rao, Stacie L. Bulfer, Yongzhao Yan, Neal Green, Barbara Mroczkowski, R. Jeffrey Neitz, Peter Wipf, Veronica Falconieri, Raymond J. Deshaies, Jacqueline L. S. Milne, Donna Huryn, Michelle Arkin, Sriram Subramaniam*

*Corresponding author. E-mail: ss1@nih.gov

Published 28 January 2016 on *Science* First Release
DOI: 10.1126/science.aad7974

This PDF file includes:

Materials and Methods

Figs. S1 to S10

Captions for movies S1 and S2

Other supplementary material for this manuscript includes the following:

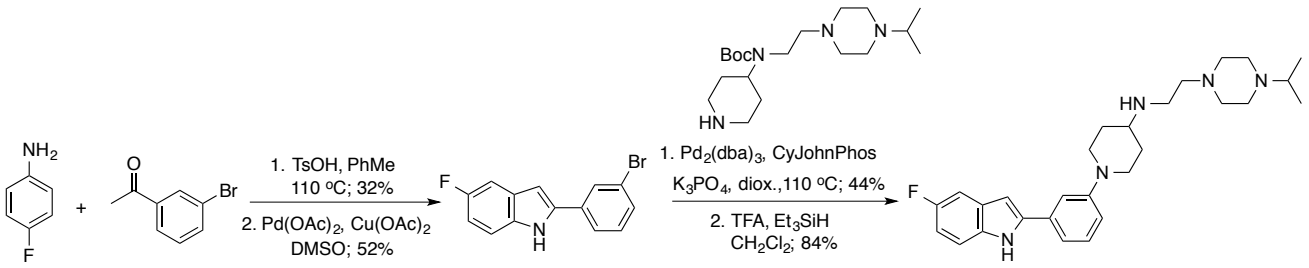
Movies S1 and S2

MATERIALS AND METHODS

Recombinant protein expression and purification

Full-length recombinant His6x-tagged human p97 was purified as described (*1*). Briefly, *E. coli* Rosetta2 (DE3) cells expressing plasmid-encoded p97 were suspended in 50 mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole, and 0.5 mM DTT. The cells were lysed, and following centrifugation, the resulting supernatant was loaded onto a Ni-NTA column and eluted with a linear gradient of imidazole. The His-tag was cleaved using TEV protease during dialysis in 25 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM DTT and the TEV protease and uncleaved p97 was removed by collecting the follow-through from a second Ni-NTA column. Tag-free p97 was purified by size-exclusion chromatography in 25 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM DTT using a Superdex 200 16/600 GL column, and stored in aliquots at -80°C without glycerol.

Synthesis of UPCDC30245



General: All reactions were performed under an argon atmosphere and all glassware was flame-dried prior to use. Reactions were monitored by TLC analysis (EM Science pre-coated silica gel 60 F_{254} plates, 250 μM layer thickness) and visualization was accomplished with a 254 nM UV light.

Purifications by chromatography were performed using SiO_2 (SiliaFlash® F60, Silicycle) or using an ISCO-Rf flash chromatography system. $^1H/^{13}C$ NMR spectra were recorded on Bruker Avance 500/125 MHz instruments. Chemical shifts are reported in parts per million with the residual solvent peak used as the internal standard ($^1H/^{13}C$: $CDCl_3$, 7.26, 77.2 ppm). Chemical shifts were tabulated as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quarter, dd = doublet of doublet, dt = doublet of triplet, m = multiplet, b = broad, app = apparent), coupling constants, and integration. All 1D NMR spectra were processed using Bruker Topspin NMR. IR spectra were obtained on an Identity IR-ATR spectrometer. LCMS-high resolution LC/MS analyses were completed on a Thermo Scientific Exactive spectrometer with purities determined using an Agilent Technologies 385-ELSD using MeCN/H_2O with 0.1% formic acid.

A solution of 3-bromoacetophenone (2.03 g, 10.2 mmol), *p*-F-aniline (1.16 g, 10.4 mmol), and TsOH•H_2O (38 mg, 0.20 mmol) in toluene (60 mL) was heated overnight under Dean-Stark conditions. The reaction mixture was cooled to room temperature, evaporated, and purified by chromatography on SiO_2 (ISCO-Rf, 0-10% EtOAc/hexanes) to give (*E*)-1-(3-bromophenyl)-*N*-(4-fluorophenyl)ethan-1-imine as a yellow oil (0.96 g, 3.3 mmol, 32%) that was used without further purification.

A solution of (*E*)-1-(3-bromophenyl)-*N*-(4-fluorophenyl)ethan-1-imine (0.853 g, 2.92 mmol), $Pd(OAc)_2$ (63.0 mg, 0.280 mmol), and $Cu(OAc)_2$ (1.42 g, 5.51 mmol) in DMSO (25 mL) was stirred at room temperature overnight (*2*). The solution was diluted with H_2O, and extracted with EtOAc. The organic layer was washed with brine, dried (Na_2SO_4), filtered and concentrated. The residue was absorbed onto SiO_2 and purified by chromatography on SiO_2 (ISCO-Rf, 0-10% EtOAc/hexanes) to give 2-(3-bromophenyl)-5-fluoro-1*H*-indole as a light yellow solid (0.19 g, 22%). The mixed fractions were

absorbed onto SiO_2 and subjected to a 2nd chromatography on SiO_2 (7/1 hexanes/Et_2O) to give additional product (0.439 g total, 1.51 mmol, 52%).

A solution of 2-(3-bromophenyl)-5-fluoro-1*H*-indole (100 mg, 0.345 mmol), K_3PO_4 (82 mg, 0.38 mmol), $Pd_2(dba)_3$ (4.6 mg, 0.005 mmol), CyJohnPhos (7.0 mg, 0.02 mmol), *tert*-butyl (2-(4-isopropylpiperazin-1-yl)ethyl)(piperidin-4-yl)carbamate[2] (0.106 g, 0.300 mmol) in anhydrous and deoxygenated dioxane (1.5 mL) in a 2-5 mL conical reaction vial was degassed by bubbling argon for 5 min. The vessel was sealed and heated at 110 °C overnight, cooled to room temperature, diluted with sat. $NaHCO_3$, and extracted with CH_2Cl_2 (3x). The organic layers were combined, washed with brine, dried (Na_2SO_4), evaporated, and purified by chromatography on SiO_2 (2% MeOH/CH_2Cl_2 and 0.1% TEA) followed by chromatography on basic Al_2O_3 (CH_2Cl_2) to give *tert*-butyl (1-(3-(5-fluoro-1*H*-indol-2-yl)phenyl)piperidin-4-yl)(2-(4-isopropyl-piperazin1-yl)ethyl)carbamate (62 mg, 0.11 mmol, 44%) as a pale yellow semi-solid.

A solution of trifluoroacetic acid (1.2 mL, 16.1 mmol) and triethylsilane (0.25 mL, 1.6 mmol) in CH_2Cl_2 (1 mL) was added to a solution of *tert*-butyl (1-(3-(5-fluoro-1*H*-indol-2-yl)phenyl)piperidin-4-yl)(2-(4-isopropyl-piperazin1-yl)ethyl)-carbamate (*3*) (88 mg, 0.16 mmol) in CH_2Cl_2 (0.5 mL). The reaction mixture was stirred under an atmosphere of N_2 at room temperature for 2 h, concentrated, diluted with sat. $NaHCO_3$, extracted with EtOAc (3x). The combined organic layer was washed with brine, dried (Na_2SO_4), filtered and concentrated. The crude residue was purified by chromatography on SiO_2 (8 to 10% MeOH/CH_2Cl_2 with 0.1% TEA) followed by chromatography on basic Al_2O_3 (0 to 10% MeOH/CH_2Cl_2) to afford 1-(3-(5-fluoro-1*H*-indol-2-yl)phenyl)-*N*-(2-(4-isopropylpiperazin-1-yl)ethyl)piperidin-4-amine (UPCDC30245) as a yellow foamy solid (61 mg, 0.13 mmol, 84%). IR

peaks (ATR, neat) 3234, 2960, 2930, 2844, 2814, 2099, 1672, 1640, 1635, 1601, 1584, 1486, 1452, 1383, 1355, 1199, 1178, 1130 cm^{-1}; ^{1}H NMR shifts (500 MHz, $CDCl_3$) δ 8.89 (s, 1 H), 7.31-7.25 (m, 3 H), 7.22 (s, 1 H), 7.10 (d, *J* = 7.5 Hz, 1 H), 6.94-6.90 (m, 2 H), 6.74 (d, J = 1.5 Hz, 1 H), 3.73 (app d, J = 12.5 Hz, 2 H), 2.86-2.77 (m, 4 H), 2.67-2.48 (m, 12 H), 2.01 (d, J = 11.0 Hz, 2 H), 1.54 (qd, J = 12.5, 2.5 Hz, 2 H), 1.06 (d, J = 6.5 Hz, 6 H); ^{13}C NMR shifts (125 MHz, $CDCl_3$) δ 158.1 (J_{CF} = 234 Hz), 152.0, 140.6, 133.3, 133.0, 129.7, 129.6 (J_{CF} = 10 Hz), 116.4, 116.1, 113.5, 111.5 (J_{CF} = 9 Hz), 110.2 (J_{CF} = 26 Hz), 105.2 (J_{CF} = 24 Hz), 99.8 (J_{CF} = 4 Hz), 58.0, 55.1, 54.5, 53.5, 48.7, 48.6, 43.4, 32.5, 18.7; ^{19}F NMR shifts (470 MHz, $CDCl_3$) -124.3; mass spectrometry (ESI) $C_{28}H_{39}N_5F$ *m/z* calculated for [M+H]+ 464.3184, found 464.3184; ELS purity (100%).

Grid preparation

Prior to grid preparation, 4 μl of a freshly thawed aliquot of purified p97 was deposited on Quantifoil R1.2/1/3 grids and plunge-frozen in liquid ethane cooled by liquid nitrogen. For experiments with added inhibitor, the protein was pre-incubated for 30 minutes at room temperature in a solution containing 0.9 mg/ml p97, 50 μM inhibitor (final 0.3% DMSO, starting from 200 μM stock solution of the compound in 1% DMSO), 1 mM $MgCl_2$ and 1 mM TCEP. For experiments with added ATPγS, the protein was pre-incubated for 30 minutes at room temperature in a solution containing 0.9 mg/ml p97, 5mM ATPγS, 1 mM $MgCl_2$ and 1 mM TCEP.

Data Acquisition

The grids were imaged using a Titan Krios transmission electron microscope (FEI Company, Hillsboro, OR) aligned for parallel illumination and operated at 300 kV, with the specimen maintained at liquid nitrogen temperatures. Images were recorded on a K2 Summit camera (Gatan, Inc., Pleasanton, CA)

operated in super-resolution counting mode, placed at the end of a GIF Quantum energy filter (Gatan, Inc.), functioning in zero-energy-loss mode with a slit width of 20 eV as described previously (*4, 5*). Data was typically collected with a defocus range between -1.0 and -2.0 μm, at a dose rate of ~2.6 e-/Å^2/s at the specimen plane), and a total exposure time of 15.2 s. Intermediate frames were recorded every 0.4 s giving an accumulated dose of ~40 e-/Å^2 and a total of 38 frames per image.

Image Processing

Movie frame alignment and CTF estimation for each micrograph were carried out as described previously (*4*). Particles were picked automatically using a Gaussian disk of 85 Å in radius as the search template. For data collected with bound inhibitor, 332,805 particles were picked from 2,068 micrographs and subjected to 8 rounds of refinement in FREALIGN v9.10 (*6*). The distribution of scores assigned to each particle by FREALIGN showed a clear bi-modal distribution and only particles in the group containing the higher scores were selected for further processing. This subset of 107,890 particles was subjected to 46 rounds of 3D classification using 3 classes. Two of the classes showed similar high-resolution features and the combined 40,913 particles that were assigned to them were subjected to an additional 8 rounds of local refinement. At this point, particles were re-extracted from the original micrographs using only the subset of frames 3-16, and used to produce a reconstruction with particle orientations derived using the full exposure followed by 8 additional rounds of local refinement in FREALIGN. A B-factor of -75 Å^2 was applied to the map for purposes of visualization.

For the unbound p97 structure, 301,685 particles were picked from 1,925 micrographs and subjected to 8 rounds of refinement. The distribution of FREALIGN scores also showed a bi-modal distribution and the 157,840 particles in the group containing higher scores were subjected to 24 rounds of 3D

classification using 3 classes. One class was discarded because it did not show high-resolution features while the other two represented "primary" and "secondary" conformations containing 69,703 and 59,492 particles, respectively (Figure S2). Particles assigned to the dominant class that showed the highest-resolution features were further classified into three sub-classes (30 additional iterations in FREALIGN). The subset of 30,616 particles assigned to the most uniform of these classes were subsequently re-extracted from the original micrographs using the range of frames 3-16, and used to produce a reconstruction with particle orientations derived using the full exposure followed by 8 additional rounds of local refinement in FREALIGN. A B-factor of -80 $Å^2$ was applied to the map.

For data collected with ATPγS-bound p97, 475,354 particles were selected from 628 micrographs. Eight rounds of 3D refinement resulted again in a bimodal distribution of scores and the 203,216 particles assigned to the lobe with the highest scores were selected and subjected to 24 rounds of classification using 5 classes. Of these, two classes were discarded because they did not show high-resolution features and had low numbers of particles. The remaining three classes represented distinct conformational states of ATPγS-bound p97: conformation I (70,330 particles), conformation II (61,909 particles), and conformation III (52,743 particles). These three sets of particles were each subjected to an additional round of 3D classification into two sub-classes (24 additional iterations in FREALIGN). In all three cases one of the two sub-classes showed better defined features and particles corresponding to the best sub-class in each set (corresponding to 33,882, 32,406, and 23,306 particles for conformations I, II and III, respectively) were re-extracted from the original micrographs using the subset of frames 3-16 and used to produce the final reconstructions with the alignments derived using the full exposure followed by 8 additional rounds of local refinement in FREALIGN. A B-factor of -85 $Å^2$ was applied to the three

maps. Refinement was also carried without imposing six-fold symmetry, resulting in the same three main classes as observed upon imposition of six-fold symmetry (Figure S9).

Structure refinement

The structure of p97 reported from earlier X-ray studies (PDB ID 3CF3) containing ADP in both nucleotide-binding regions was used as a rough starting model to initiate refinement. Using the program Chimera (*7*), the N, D1 and D2 domains were placed separately into the map to get this initial structure and the atomic model was rebuilt manually for the entire polypeptide as described below. For the UPCDC30245-bound structure, the inhibitor was first manually docked into the additional density. Subsequently, the entire structure was refined further in the same way as that of native p97 including real-space refinement using the programs COOT (*8*) and PHENIX (*9*). The fit of the model to the map was analyzed and then subjected to reciprocal space refinement using PHENIX. Rigid-body fitting was used to refine the individual domains, followed by several rounds of positional and isotropic B-factor refinement and geometry minimization. Six-fold symmetry was imposed during refinement steps and solvent molecules were added at the end. The resulting atomic model was examined in COOT and manual changes were made to selected residues as needed, guided by the density map. The progress of refinement was monitored by the analysis of the results using the program Molprobity (*10*).

The initial models for refinement of p97 structures in the presence of ATPγS and intrinsically bound ADP were based on the identity of the nucleotides present in the various nucleotide-binding regions of the D1 and D2 domains. For the refinement of the conformational states (I and II) that demonstrated either the presence of ADP molecules in both domains, or ADP in D1 and ATPγS in D2, the high-resolution p97 structure with ADP bound was used as the starting model. The respective nucleotides

were docked manually into density. The refinement strategy was identical to that described above for native p97. For the conformational state that exhibited the presence of ATPγS in both domains, an initial model for refinement was created by combining the amino acid residues corresponding to D1 (12-462) from PDB ID 3HU1, and D2 (463-763) from PDB ID 3CF3. This hybrid model was initially fitted by sequential rigid-body refinement that individually refined the spatial fits of the two domains independently. Thereafter, the ATPγS molecules were placed into density in their respective binding pockets and subjected to local real-space refinement.

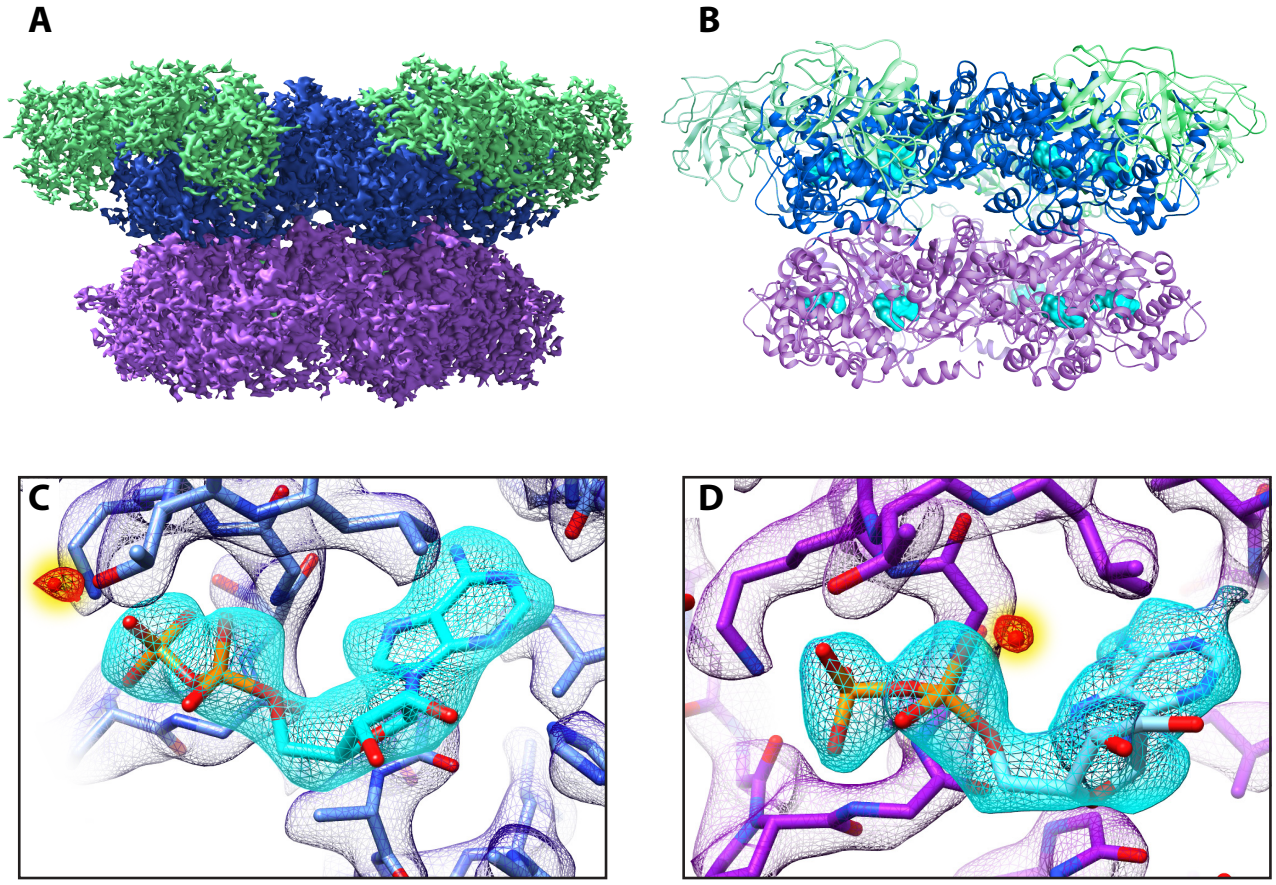


Figure S1
Structure of full-length p97 at 2.4 Å resolution. (A) Cryo-EM derived density map with the N (green), D1 (blue) and D2 (purple) domains highlighted. (B) Ribbon model of cryo-EM derived structure of p97, color-coded as in (A). (C, D) Density for bound ADP in D1 and D2 domains, respectively (shown in cyan). Water molecule (red) highlighted in yellow color.

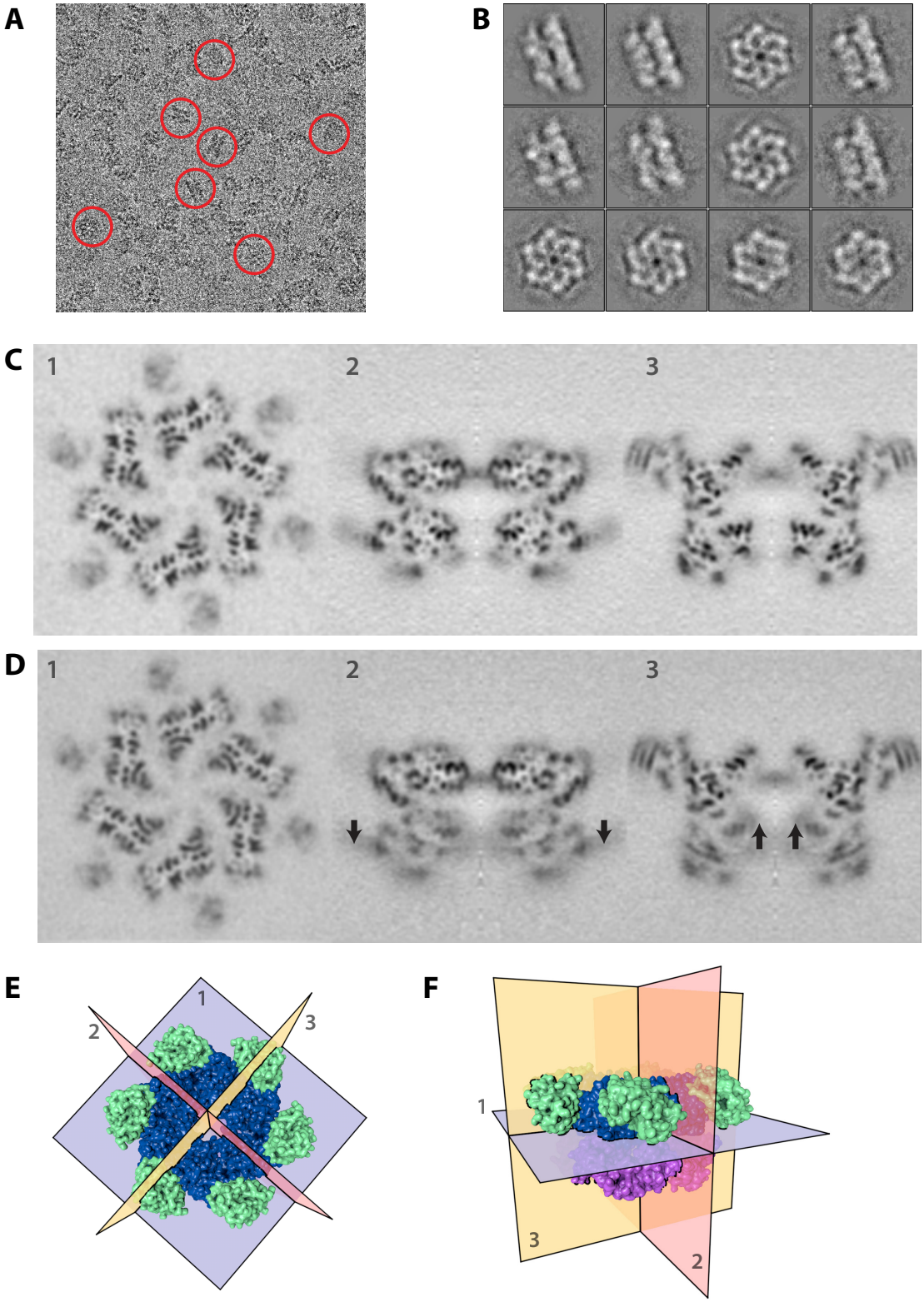


Figure S2

Figure S2
Intermediate stages in structure determination of wild-type p97. (A) Raw cryo-EM image displaying p97 hexamers in different orientations; several representative molecular images are circled. (B) Reference-free 2D class averages showing that both top and side views are represented in the data. (C, D) Sections through cryo-EM density map of native p97 for the primary (C) and secondary (D) classes. The first section is a horizontal slice through the lower D1 domain, while the second and third sections are orthogonal cuts through the center of the map showing side views. (E, F) Perspective views of molecular surface models showing the locations of the three planes used to produce the sectional images shown in panels (C) and (D). The maps show that the D1 and N domains display similar structures in the two conformational classes, but the density of the D2 domain is more diffuse in the class shown in (D), consistent with greater flexibility. The arrows in (D) indicate the locations of subtle differences in conformation of the D2 domain between the two classes. Slices in this figure were generated from uncorrected maps, and the images enable identification of both ordered and fuzzy secondary structural elements in the two classes. A total of 69,703 and 59,492 particles were distributed into the primary and secondary classes, respectively. A subset of the particles that contributed to the dominant class (30,616 out of 69,703 particles) was used to produce the map of native p97 at 2.4 Å resolution shown in Figure S1. The minor class produced a map at lower resolution (~ 3.5 Å).

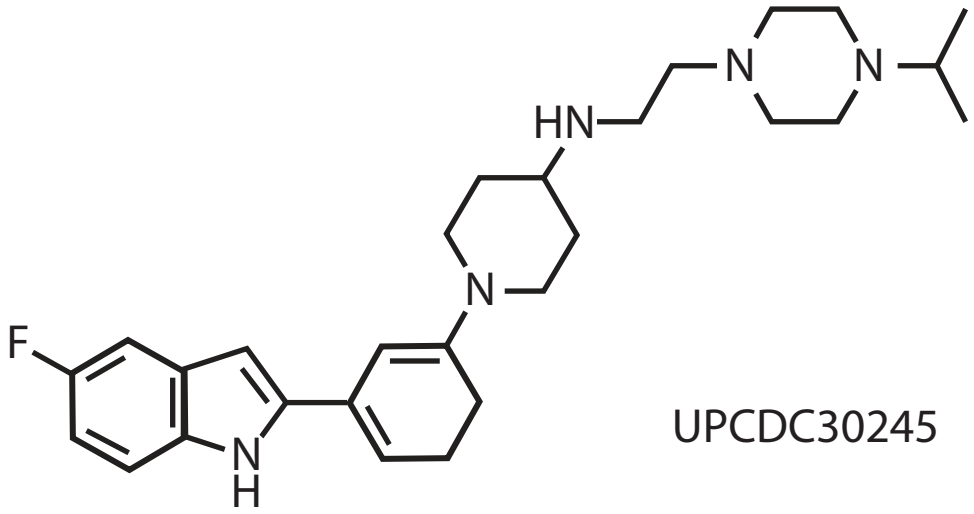


Figure S3
Chemical structure of the inhibitor UPCDC30245 used in this study.

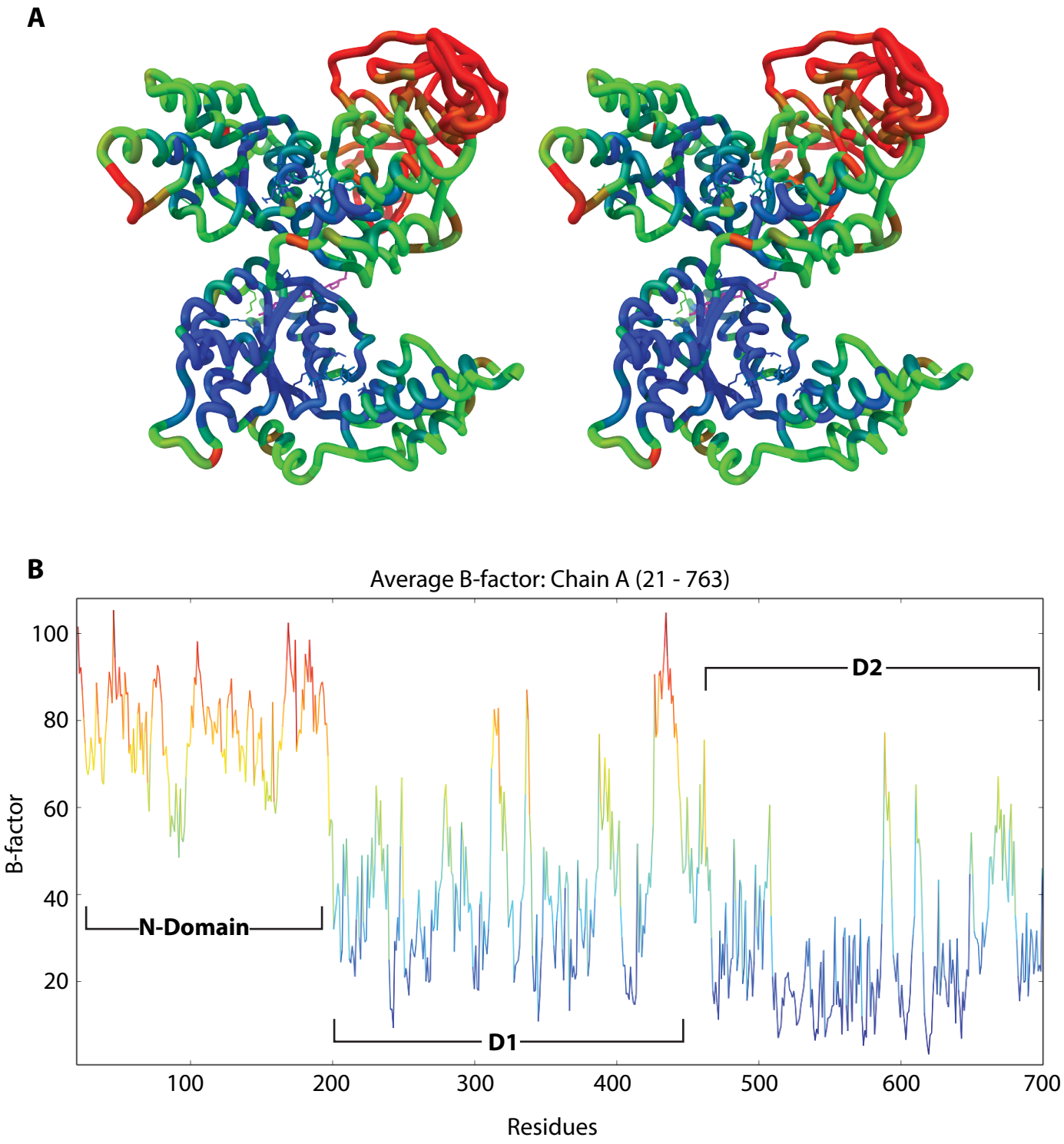


Figure S4
B-factor analysis of the structure of p97 bound to UPCDC30245. (A) Stereo view of the p97 protomer, color-coded to indicate variation in B-factors in the refined structure from low (blue) to high (red). (B) Graphical representation of the average B-factors plotted against the protein sequence. The amino acid residues encompassing the various domains are indicated.

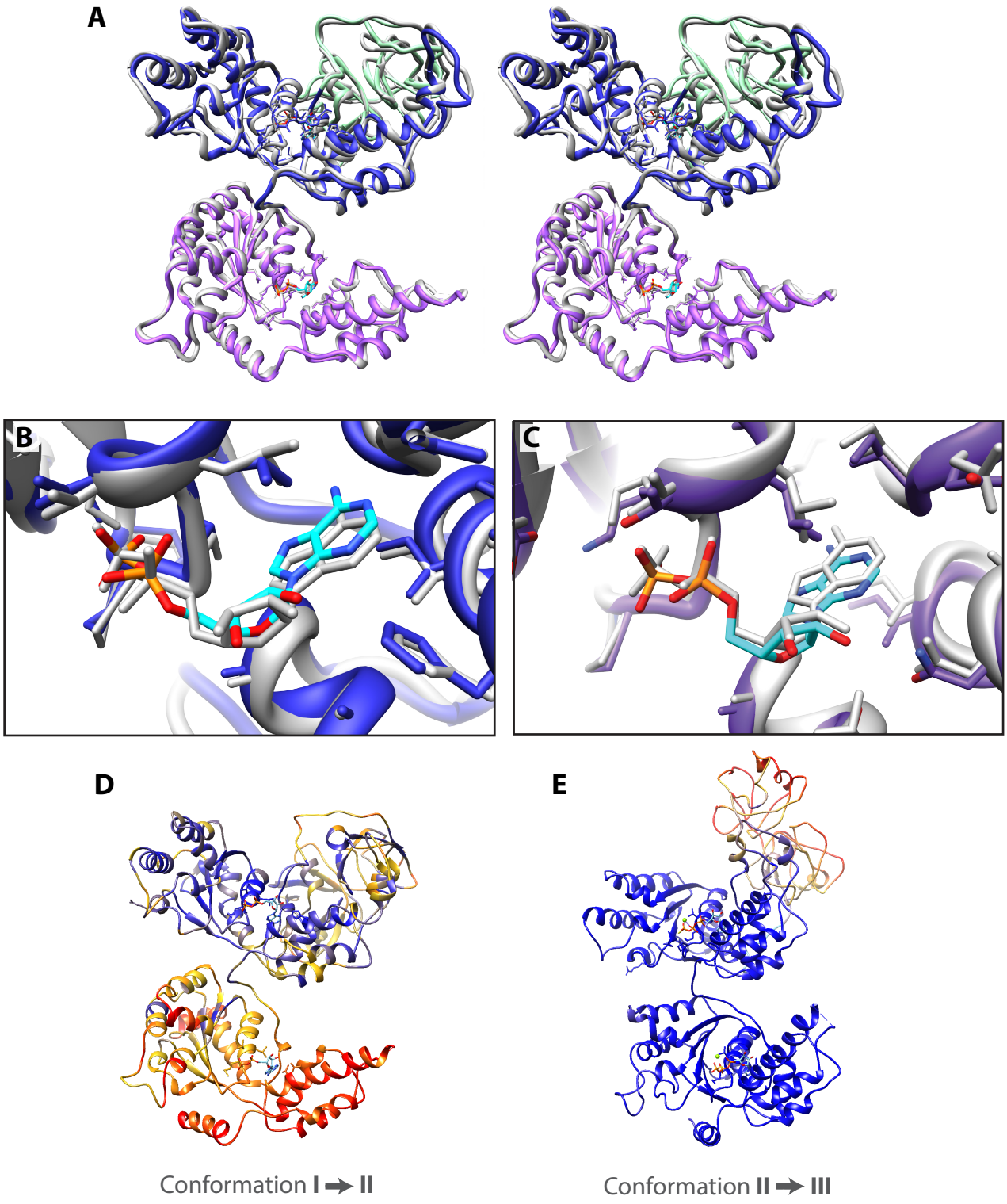


Figure S5
Comparison of the structures of p97 in the absence and presence of the UPCDC30245 inhibitor. (A) Stereo view of the p97 protomer shown in ribbon representation, with the N, D1 and D2 domains of native p97 color coded as in Figure 1, and the inhibitor-bound structure shown in grey. The two structures are very similar with an overall main chain RMSD of 0.84 Å. (B, C) Comparison of the ADP binding sites shows that there are only minimal differences between the two structures even at the nucleotide binding sites. As in (A), the inhibitor-bound structure is colored in grey. (D, E) Changes between conformations I to II (D) and from II to III (E), color-coded from blue to red to reflect the extent of change in structure.

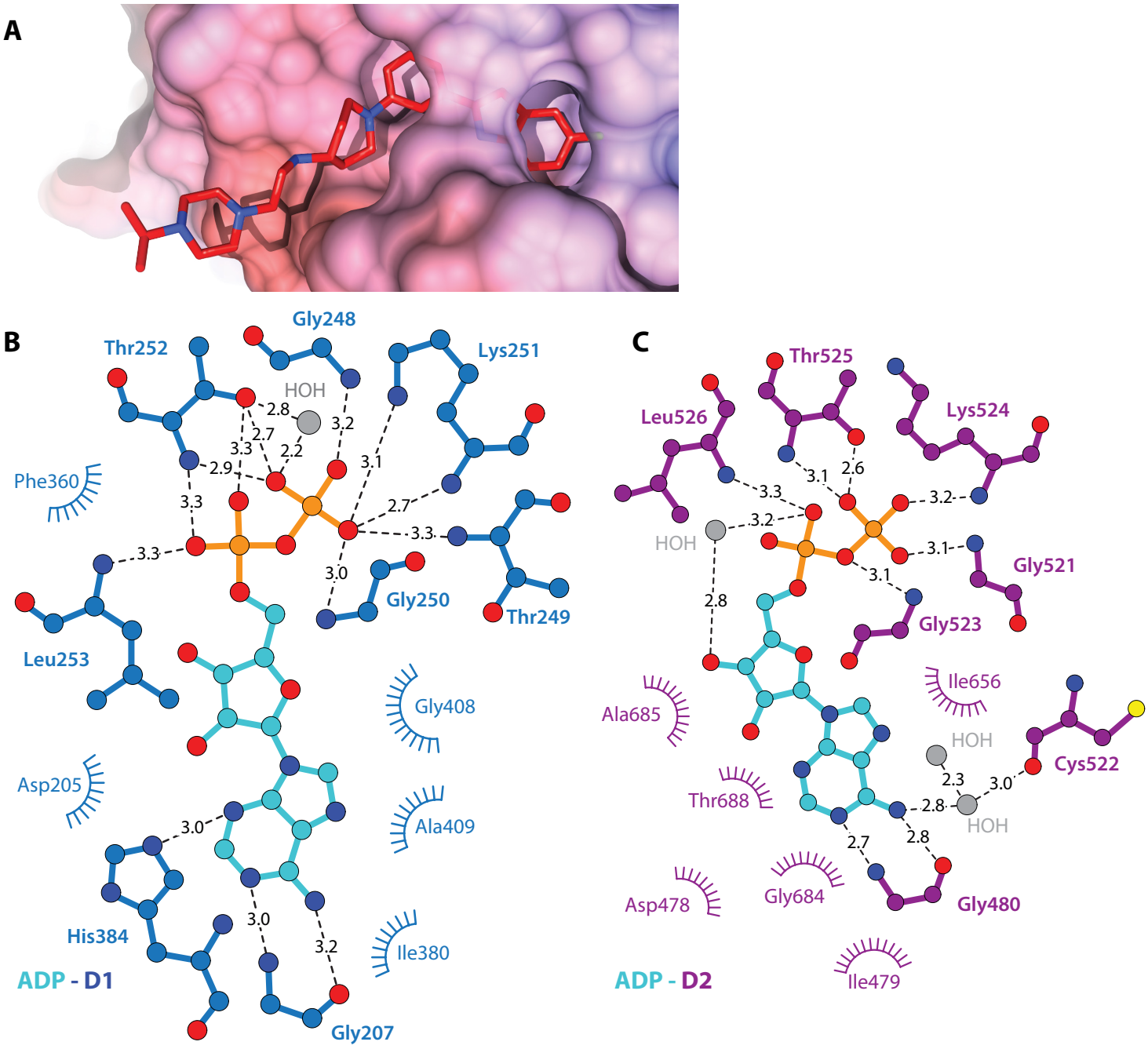


Figure S6

Inhibitor and nucleotide binding sites. (A) Electrostatic surface representation of the inhibitor binding pocket. The figure illustrates the cavity into which the phenyl indole end of the inhibitor inserts and the charge complementarity on the outer edge which presumably helps to stabilize binding of the inhibitor. (B, C) LIGPLOT renderings of the residue environments of ADP bound in D1 and D2 binding sites, respectively. The plots identify all residues in the immediate vicinity of the bound nucleotides and show the relevant H-bond distances.

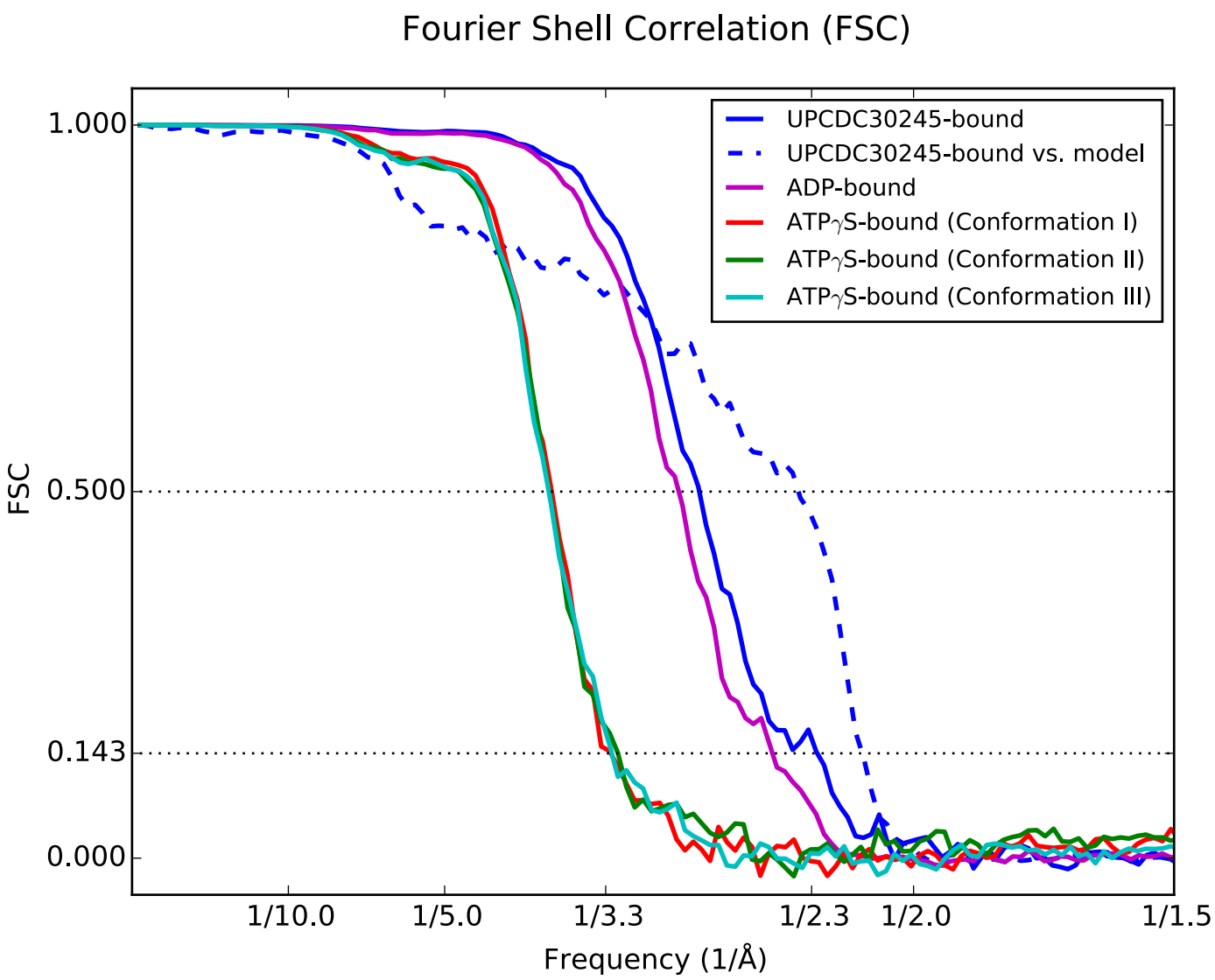


Figure S7
Fourier Shell Correlation (FSC) plots to measure resolution of the cryo-EM maps obtained for p97 in the native state, in complex with UPCDC30245, and with ATP S (three conformations). Shown in blue are the FSC curve between two semi-independently refined halves of inhibitor-bound p97 (solid blue) and FSC curve calculated between the map shown in Figs. 1-2 and the map computed from the cryo-EM-derived atomic model (dashed blue), both showing a resolution value of 2.3 Å at the 0.143 and 0.5 cutoffs, respectively. FSC curves between two semi-independently refined halves are shown for native ADP-bound p97 (magenta) showing a resolution of 2.4 Å, and for the three conformational states in the presence of ATPγS showing resolutions of 3.3 Å (red), 3.2 Å (dark green) and 3.3 Å (pale green) for conformational states I, II and III, respectively.

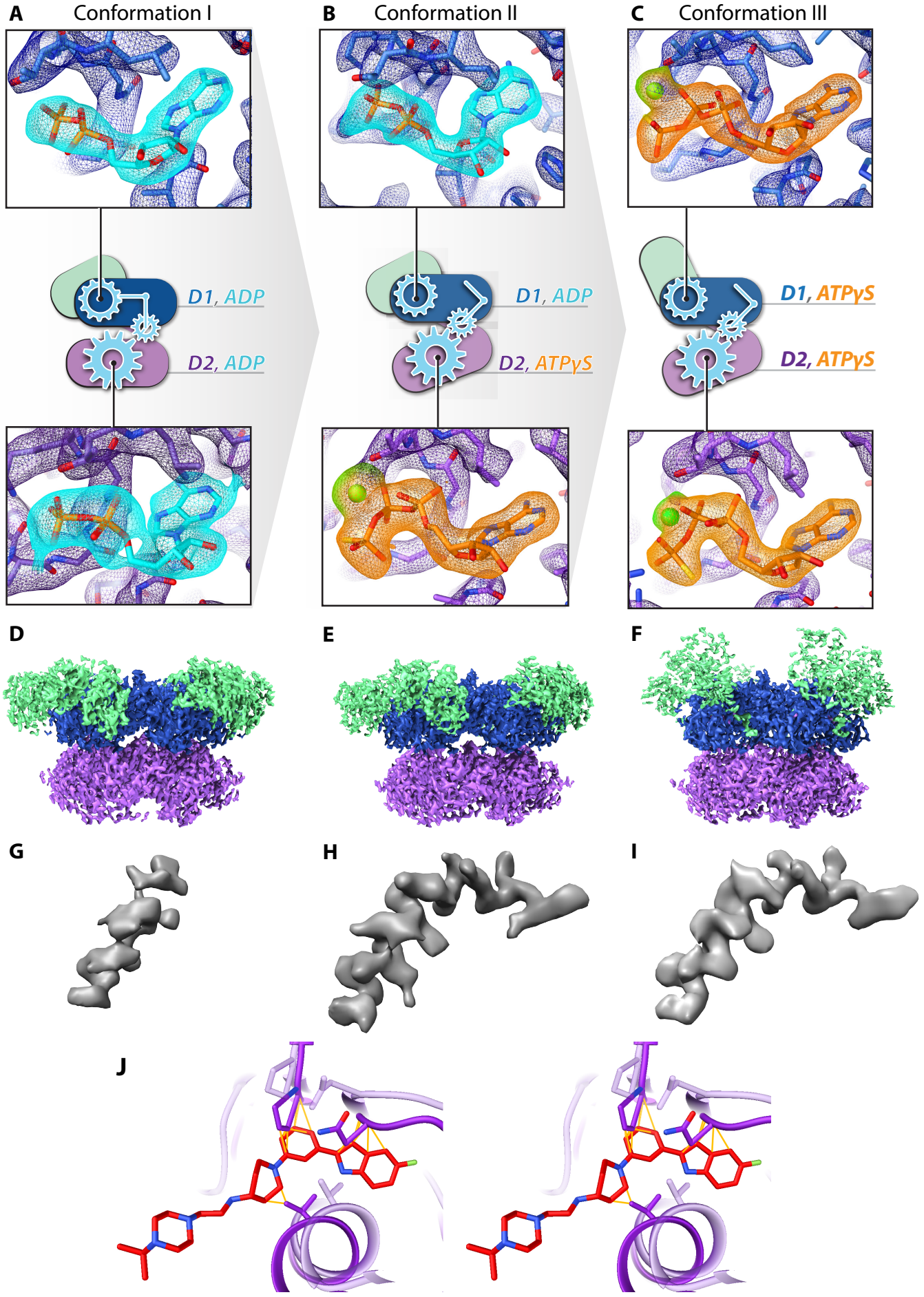

A
Conformation I
B
Conformation II
C
Conformation III
D1, ADP
D2, ADP
D1, ADP
D2, ATPγS
D1, ATPγS
D2, ATPγS
D
E
F
G
H
I
J

Figure S8

Figure S8

Visualization of densities for nucleotides in the three conformational states. (A-C) Densities for nucleotides occupying the D1 and D2 domains enable unambiguous assignment to ADP or ATPγS. In conformation I, ADP occupies both nucleotide-binding sites. In conformation II, ADP is bound to D1, while ATPγS is bound to D2. In conformation III, ATPγS is bound to both D1 and D2 domains. The schematic diagram at the center of each panel illustrates the regions that undergo conformational changes in the sequential transition from conformation I to III. (D-F) Overview of the density maps for each conformation, colored as in Figure 1. (G-I) Stabilization of the C-terminal helix of the D2 domain. In the presence of ADP alone (conformation I), the model ends at residue 763. In the presence of ATPγS in conformations II and III, the stabilization results in traceable density leading to an extension of the model to residue 768. (J) Clash between conformation II and inhibitor binding. Stereo view of the binding site of UPCDC30245 in conformation I, superposed with conformation II to show that the latter conformation is incompatible with binding inhibitor in this position.

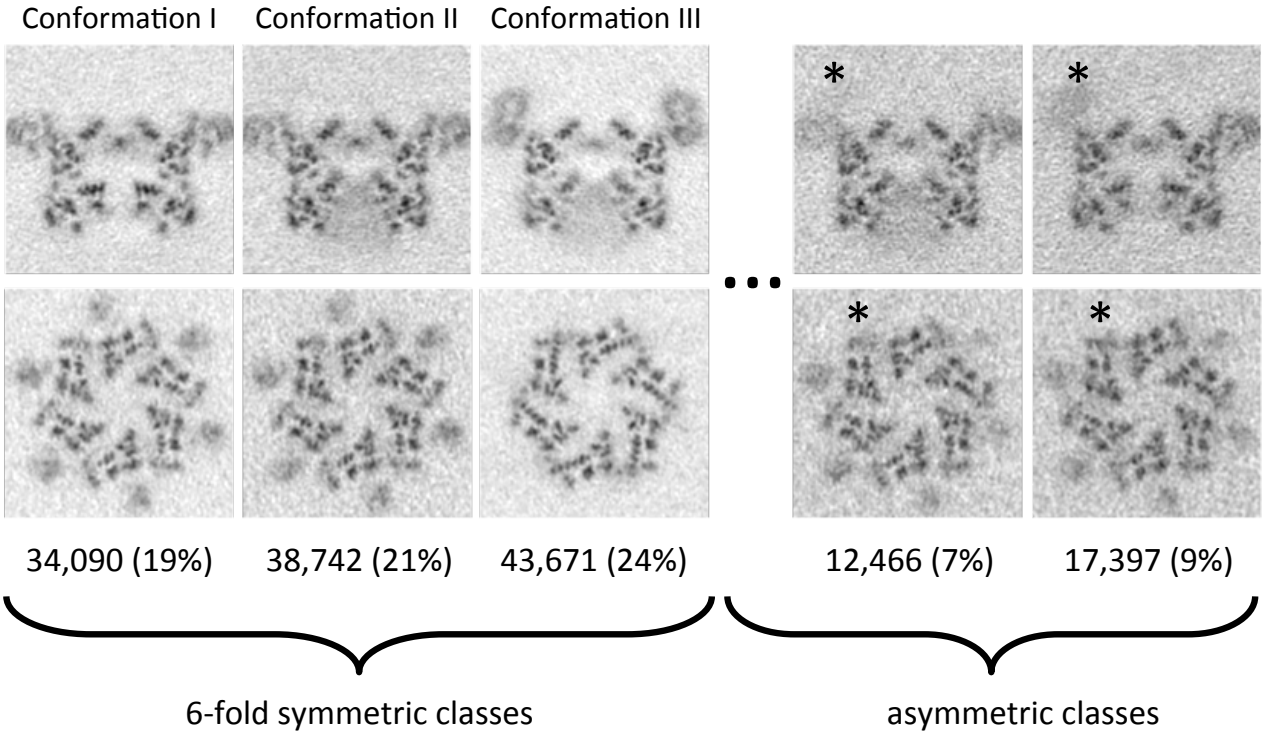


Figure S9
Sections through cryo-EM density maps of ATPγS -bound p97 conformations refined without symmetry imposition. The three major classes that emerged at the end of the refinement correspond to symmetric structures that are the same as those obtained when six-fold symmetry was imposed during refinement. Two minor classes (~ 7% and ~ 9%) are also observed in which the N-domain is displaced upwards in only some of the protomers (indicated by asterisks). The low number of particles assigned to these asymmetric classes resulted in resolutions that were not high enough to allow determination of the stoichiometry or identity of ligands bound to each monomer. Side and top views of central slices through the map are shown for each conformation, with the slices obtained as indicated in Figure S2. Numbers of particles assigned to each conformation are indicated in each case. This analysis shows that although there is clearly a propensity for asymmetric states, the three symmetric states explain the predominant source of variability within the particle population. The fact that we can determine the structures of each state to near-atomic resolution indicates that there is sufficient similarity between the units of each hexamer, at least to a resolution of 3.2 Å.

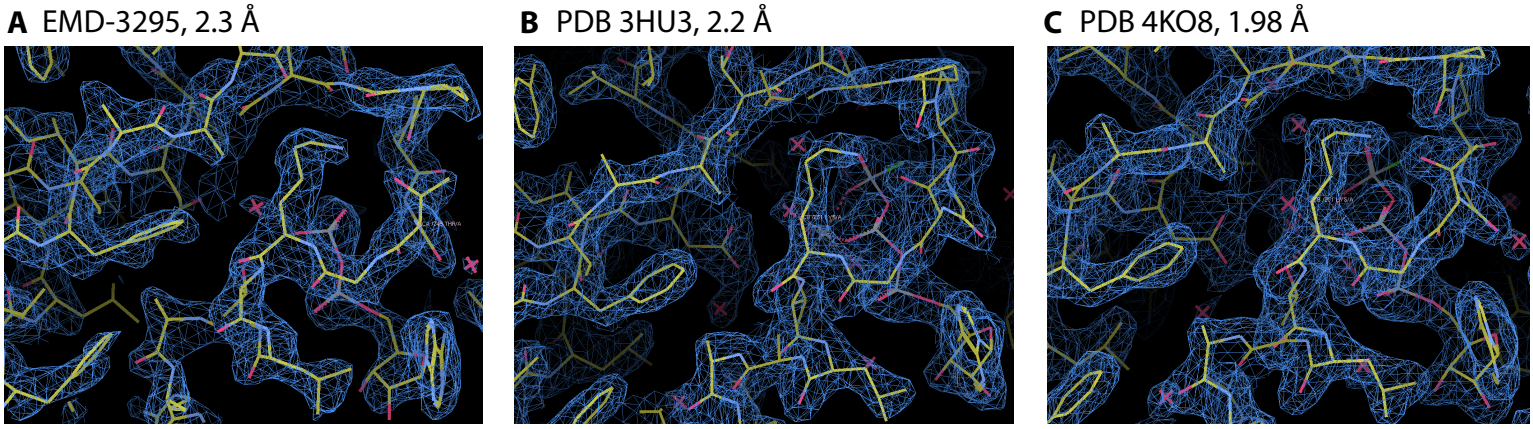


Figure S10
Comparison of 2.3 Å cryo-EM density map with X-ray maps of p97 fragments using X-ray crystallography of similar resolution. A) Density map in the vicinity of Lys 251 of the 2.3 Å resolution UPCDC30245 bound p97 cryo-EM structure reported here. B) and C) 2Fo-Fc maps of the same region from previously published X-ray structures of the N-D1 fragment at 2.2 Å resolution (PDB ID: 3HU3) and at 1.98 Å resolution (PDB ID: 4KO8), respectively.

Supplementary Movie S1: Views from selected regions of the density map of UPCDC30245-bound p97.

Supplementary Movie S2: Density maps of the three conformational states in the region close to the cytoplasmic face of the central pore.

REFERENCES

1. T. F. Chou *et al*., Specific inhibition of p97/VCP ATPase and kinetic analysis demonstrate interaction between D1 and D2 ATPase domains. *J Mol Biol* **426**, 2886-2899 (2014).
2. Y. Wei, I. Deb, N. Yoshikai, Palladium-catalyzed aerobic oxidative cyclization of N-aryl imines: indole synthesis from anilines and ketones. *J Am Chem Soc* **134**, 9098-9101 (2012).
3. C. Alverez *et al*., Structure–activity study of bioisosteric trifluoromethyl and pentafluorosulfanyl indole inhibitors of the AAA ATPase p97. *ACS Med. Chem. Lett*., (2015).
4. A. Bartesaghi, D. Matthies, S. Banerjee, A. Merk, S. Subramaniam, Structure of beta-galactosidase at 3.2-A resolution obtained by cryo-electron microscopy. *Proc Natl Acad Sci U S A* **111**, 11709-11714 (2014).
5. A. Bartesaghi *et al*., 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147-1151 (2015).
6. D. Lyumkis, A. F. Brilot, D. L. Theobald, N. Grigorieff, Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol* **183**, 377-388 (2013).
7. E. F. Pettersen *et al*., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
8. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
9. P. H. Zwart *et al*., Automated structure solution with the PHENIX suite. *Methods Mol Biol* **426**, 419-435 (2008).
10. V. B. Chen *et al*., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).