Supporting Information

Structure-based design of selective LONP1 inhibitors for probing *in vitro* biology

Laura Kingsley,[*,†] *Xiaohui He, Matthew McNeill, John Nelson, Victor Nikulin, Zhiwei Ma, Wenshuo Lu, Vicki W. Zhou, Mari Manuia, Andreas Kreusch, Mu-Yun Gao, Darbi Witmer, Mei-Ting Vaillancourt, Min Lu, Sarah Greenblatt, Christian Lee,*[‡] *Ajay Vashisht, Steven Bender,*[§] *Glen Spraggon, Pierre-Yves Michellys, Yong Jia, Jacob R. Haling,*[||] *Gérald Lelais**

Genomics Institute of the Novartis Research Foundation, 10675 John J. Hopkins Dr., San Diego, CA, 92121.

CONTENT

METHODS

X-RAY CRYSTALLOGRAPHY

Crystallization experiments were carried out with LONP1. Protein was thawed and diluted with 20 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM TCEP to a final protein concentration of 15 mg/ml. For the co-crystal structures, 1 mM compound (bortezomib, **9a**, **12d**) was added along with 1 mM $MgCl_2$; for the apo structure, no additional reagents were added. All solutions were incubated overnight at 4 ºC.

Crystallization was done using sitting drop vapor diffusion using the JCSG coarse screens[1] and were all grown at 20 ºC. Multiple crystals were harvested and cryo-protection was accomplished by adding glycerol to the drop prior to looping (see below for details on individual crystals). All crystals were flash cooled in liquid nitrogen at -180 ºC prior to data collection from single crystals at Beamline 5.0.3 of the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory in Berkeley, CA.

Data reduction and structure solution were done using the software packages PHENIX,[2] HKL2000[3] and MOSFLM.[4] Initial phase determination was done with molecular replacement programs MOLREP[5] and PHASER[6] from the CCP4 program package[7] with the first "apo" structure solved using a probe molecule of the coordinates of the pdb code 2X36.[8] Structure refinement for all structures was carried out using iterative model build cycles with BUSTER,[9] PHENIX[2] and COOT[10] until the final model quality (evaluated using tools from the PHENIX package) and agreement with experimental data converged, where appropriate non-crystallographic LSSR NCS coordinate restraints were used as implemented in BUSTER.[9]

The coordinates and structure factors have been deposited in Protein Data Bank with accession codes 6WYS (LONP1 apo form), 6X27 (bortezomib-bound), 6WZV (**9a**-bound) and 6X1M (**12d**-bound), respectively. Details of the crystallography are reported in **Table S1** below.

MD SIMULATIONS

The 20S proteasome apo structure 5LE5 was used for apo simulations. Each of the proteolytically active subunits were extracted from the structure (β5, β2, β1). Homologous subunits were compared to one another (*e.g.*, β5 chain K was compared to β5 chain Y), and all subunits were found to contain only minor discrepancies (RMSDs; β1= 0.047, β2= 0.057, β5=0.053), therefore chains were not simulated independently. Chains K, H, and N, were selected as starting points for β5, β2, and β1 subunits, respectively. For LONP1, the in-house apo structure of the protease domain-only was used (PDB ID 6WYS).

For the holo simulations containing bortezomib, the structure 5LF3 and the in-house structure 6X27 were used for the 20S proteaseome and LONP1, respectively. As with the apo form of the 20S proteasome, the chains were internally compared and found to be nearly identical. Chains K, H, and N, were selected to as starting points for β5, β2, and β1 subunits, respectively. The C-terminal tail of the β2 subunit was removed at T197. To simplify the simulations, the reactive portion of bortezomib was removed by truncating the bond connecting the boron to the main chain carbon as has been described by Felix et.al.[11] For holo simulations of LONP1 containing bortezomib derivatives, the same truncation procedure/preparation procedure was followed.

In all systems, water molecules within 5 Å of the chains were retained but other heteroatoms, ions, and waters further than 5 Å were removed. Hydrogen atoms and histidine states were assigned using Maestro from Schrödinger and were manually confirmed. Missing residues were

added using Maestro and appropriate rotamers were manually selected. Proteins were solvated in a 10 Å box using explicit waters (TIP3P water model) and neutralized as follows using Cl or Na atoms.

Simulation parameters. To simulate the covalent bond formed between the boronic acid of the inhibitor and the S855 of LONP1 or the T1 of 20S proteasome, a harmonic restraint was employed. The restraint was created between the reactive oxygen in the enzyme and the carbon atom adjacent to the removed boronic group. The restraint was set at 2.5 Å with a threshold of 2.0-4.0 Å.

MD simulations were performed using Amber 16 with the ff14SB force field. Systems were minimized using Cartesian restraints for 1000 steps using steepest decent for the first 500 steps followed by 500 steps using the conjugate gradient algorithm. Next, a non-restrained 2500 step minimization was performed (first 1000 steps using steepest decent followed by 1500 steps using the conjugate gradient algorithm). Next, the system was heated from 0 to 300 K over 20 ps at constant volume (NVT) using the SHAKE protocol to fix hydrogen atoms. This was followed by a pressure equilibration (NPT) for 5 ns. Production runs were then carried out for 100 ns on each system. All simulations were performed using the cuda-enabled pmemd protocol on either the Nvidia Titan X or Tesla P100. Simulations were performed in duplicate.

Analysis. To calculate the RMSF of the binding site residues, the trajectories were aligned by selected binding site residues (residues 11-65 for the 20S proteasome and 765-782, 842-859 for LONP1). Next, residues within 5 Å of bortezomib in the respective crystal structure were selected and the RMSF of the residue was calculated over the length of the trajectory using the cpptraj module of Amber. The binding pocket size was calculated using the icmPocketFinder tool within MolSoft. The tolerance of 5.0, a slight increase from the default value of 4.6, to

prevent offshoots into the lower regions of the pocket that were occasionally observed. This was automated by only reporting the pocket(s) that occurred within 3.0 Å of the reactive Ser/Thr.

DOCKING

The covalent docking module in MolSoft ICM was used to perform all docking simulations. All docking runs were performed in triplicate and the best scored pose (lowest numerical value) was selected for each docking model. A total of four docking models were tested for each protease subunit; the apo crystal structure, the bortezomib-bound crystal structure, and two "enlarged pocket" models.

The "enlarged pocket" models for docking were selected from the MD simulations. Frames from the MD trajectories were first clustered using average linkage clustering within the cpptraj package of AmberTools and clusters containing less than 50 members were discarded. The pocket size of each cluster centroid was calculated and the structure with the largest pocket was selected as the "enlarged pocket" model for each system.

Compounds were docked to each model and poses were ranked using the default internal ICM scoring function, where the smaller the number the better the pose. The performance of each model was evaluated using the co-crystalized pose of bortezomib. Models that failed to generate a bortezomib pose less than 3.0 Å from the crystal pose were eliminated. This amounted to 6 models in total; all four apo-derived enlarged pocket models and the apo structures of both LONP1 and the β5 subunit (these models are indicated with a * in **Table S2**).

Next, poses were manually grouped into pose types; bortezomib-like, **9a**-like, or alternative. Poses in which the P1 substituent was not in the P1 pocket/groove were automatically placed into the alternative grouping (**Table S2**). For poses where the P1 substituent was found in the P1

pocket/groove, the P2 and P3 had to be roughly in the same orientation as either bortezomib or **9a** to be fitted classified into either of those groups, otherwise they were classified as alternative.

The lowest (best) scored pose that was either in the bortezomib-like or **9a**-like conformation was selected as the final model/pose for each compound:protease pair. Compounds for which only alternative poses were identified or for which the best score was >-10.0 were excluded (*e.g.*, **11b** in the 20S β2 subunit). **Table S2** contains the scores for all models. Lastly, poses were manually inspected and three "top-scoring" poses were replaced with the second-best scoring model based on chemical intuition. Those compounds were **14**: LONP1, **9g**:β1 and **9c**:β2. In all three cases, the top ranked pose was bortezomib-like, but the **9a**-like pose was expected based on the molecular chirality. Closer inspection revealed that all of these had a close-second ranked pose in the **9a**-like conformation and therefore this pose was selected over the bortezomib-like pose.

The resulting poses and docking models are thought to be fairly robust. The β2 subunit performed the worst both in terms of overall scores and the number of compounds for which no pose was identified. This is in agreement with literature suggesting that bortezomib and bortezomib-analogs do not bind to the β2 subunit in physiologically relevant conditions.[12]

Notably for all four protease sites, binding conformations with reasonable scores were achievable for nearly all compounds (Supplemental **Table S2**). Compounds **12a-g** and **14** scored about 1.5x worse in the β5 pocket of the 20S proteasome compared to LONP1, which is in agreement with experimental data.

LONP1-3XFLAG STABLE CELL LINES

Flp-In™ T-REx™-293 cells were obtained from Thermo Scientific. These cells were cultured and maintained as described by the manufacturer. Flp-In™ T-REx™-293 cells stably expressing LONP1-3X-FLAG construct were generated using the Flp-In system (Invitrogen) according to manufacturer's instructions. The expression of LONP1-3X-FLAG in stable Flp-In™ T-REx™-293 cells was driven by adding doxycycline overnight to a final concentration of 500 ng/mL.

AFFINITY PURIFICATION

Six 15 cm tissue cultures plates of LONP1-3XFLAG expressing Flp-In™ TREx™-293 cells were grown control parental cell lines. Each set of cells were either treated with DMSO or with LONP1 inhibitor **14**. The control cells were harvested and lysed in IP buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM EDTA, 5% glycerol, 0.1% NP-40, 1 mM DTT, 0.5 mM PMSF, 1 μM pepstatin, 1 μM leupeptin and 2 μg/mL aprotinin). Clarified and normalized protein lysate was incubated with 100 μL of equilibrated anti-FLAG M2 agarose beads (Sigma) for 2 h at 4°C. The beads were washed four times with 1 mL of IP buffer per wash before eluting with two sequential elution steps of 500 μL using FLAG elution buffer (IP buffer without NP-40 supplemented with 250 μg/mL of 3xFLAG peptide (Sigma)). Eluted protein complexes fractions were pooled and precipitated by the addition of trichloroacetic acid (TCA) to a final concentration of 20% and incubated on ice for 60 min. The proteins were precipitated by centrifugation at 16,000 xg for 30 min. The final precipitate was washed twice with acetone and air dried prior to further analysis.

MASS SPECTROMETRY ANALYSIS

The precipitates were dissolved in digestion buffer (8 M Urea in 100 mM Tris pH 8.5), reduced, alkylated and digested. The desalted peptide digests were analyzed and fractionated online using a 75 µM inner diameter fritted fused silica capillary column with a 5 µM pulled electrospray tip and packed in-house about 18 cm long with 3 µM reversed phase particles (ReproSil-Pur C18-AQ - Dr. Maisch GmbH HPLC). The samples were first loaded onto a trap column packed with 3 µM reversed phase particles (ReproSil-Pur C18-AQ - Dr. Maisch GmbH HPLC) of 2.5 cm length. An easy-nLC 1200 ultra-high pressure liquid chromatography (HPLC) system (Thermo Scientific) was used to deliver the linear acetonitrile gradient with buffer A (0.1% formic acid water) and buffer B (0.1% formic acid water, 80% MeCN) starting from 4% buffer B to 35% over 80 min at a flow rate of 200 nL/min, followed by a 10 min ramping up to 80% acetonitrile and a 5 min hold at 80% buffer B. The column was re-equilibrated with 2% buffer B for 2 min before next run.

An Orbitrap Fusion Lumos (Thermo Scientific) mass spectrometer was used for MS/MS analysis. Survey scans for peptide precursor were performed from 375 to 1500 m/z at 60 K FWHM resolution (at 200 m/z) with AGC target value of $7X10^5$ and maximum injection time of 50 ms. The instrument was set to run with 3 s cycle time for the survey and MS/MS scan. After each survey scan, tandem MS was done on most intense precursor with charge state from 2 to 7. HCD fragmentation was done with 25% collision energy and resulting fragments were detected in orbitrap with 30 K resolution. The AGC target for MS/MS was set to $5X10^4$ with maximum injection time of 100 ms. The dynamic exclusion was set to 60 s with a 10 ppm mass tolerance around precursor and its isotopes.

Raw data was processed using Thermo Scientific Proteome Discoverer software version 2.4.0.305. MS/MS spectra were searched with SEQUEST HT using a reviewed human uniprot

dataset. The enzyme specificity was selected for trypsin (full), allowing for up to two missed cleavages. Carbamidomethylation (+57.021 Da) was chosen as fixed modification on cysteine residues. Methionine oxidation (+15.9949 Da) and acetylation of the protein N-terminus (+42.0106 Da) was selected as variable modification. Precursor mass tolerance was set to 20 ppm and fragment ion were searched at 0.1 Da tolerance. Peptide spectral matches (PSM) were validated with percolator algorithm[13] based on 1% FDR q-values. With proteome discoverer software peptide identification grouped into proteins according to law of parsimony and filtered using 1% FDR. Minora feature detector (the precursor ion quantifier node) warranted a minimum trace length of 5. In order to calculate precursor ion intensities, feature mapper was set "true" for RT alignment requiring mass tolerance of 10 ppm and abundance of precursor ion was quantified from intensity and level of peptide identification (1% FDR).

REFERENCES

1. Page, R.; Grzechnik, S. K.; Canaves, J. M.; Spraggon, G.; Kreusch, A.; Kuhn, P.; Stevens, R. C.; Lesley, S. A. Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Cryst.* **2003**, D59, 1028-1037.

2. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **2010**, D66, 213-221.

3. Otwinowski, Z.; Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. In *Methods Enzymol.*, Academic Press: 1997; Vol. 276, pp 307-326.

4. Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst.* **2011**, D67, 271-281.

5. Vagin, A.; Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **1997**, 30, 1022-1025.

6. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Cryst.* **2007**, 40, 658-674.

7. Potterton, E.; McNicholas, S.; Krissinel, E.; Cowtan, K.; Noble, M. The CCP4 molecular-graphics project. *Acta Cryst.* **2002**, D58, 1955-1957.

8. Garcia-Nafria, J.; Ondrovicová, G.; Blagova, E.; Levdikov, V. M.; Bauer, J. A.; Suzuki, C. K.; Kutejová, E.; Wilkinson, A. J.; Wilson, K. S. Structure of the catalytic domain of the human mitochondrial Lon protease: proposed relation of oligomer formation and activity. *Prot. Sci.* **2010**, 19, 987-999.

9. Smart, O. S.; Womack, T. O.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Cryst.* **2012**, D68, 368-380.

10. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004**, D60, 2126-2132.

11. Felix, J.; Weinhäupl, K.; Chipot, C.; Dehez, F.; Hessel, A.; Gauto, D. F.; Morlot, C.; Abian, O.; Gutsche, I.; Velazquez-Campoy, A.; Schanda, P.; Fraga, H. Mechanism of the

allosteric activation of the ClpP protease machinery by substrates and active-site inhibitors. *Sci. Adv.* **2019**, *5*, eaaw3818.

12. Groll, M.; Berkers, C. R.; Ploegh, H. L.; Ovaa, H. Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome. *Structure* **2006**, 14, 451-456.

13. Käll, L.; Canterbury, J. D.; Weston, J.; Stafford Noble, W.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* **2007**, *4*, 923-925.

SUPPORTING TABLES

Table S1: Human LONP1 protease domain crystallography.

Table S2. Docking models and scores.

Table S3. Molecular formula strings.

Table S1: Human LONP1 protease domain crystallography.

	Apo (PDB 6WYS)	**Bortezomib (PDB 6X27)**	**9a (PDB 6WZV)**	**12d (PDB 6X1M)**
Data Collection (highest res shell)				
Crystallization	1.6 M $(NH_4)_2SO_4$ 0.1M NaCl 0.1 M HEPES pH 7.5	5% PEG-6000, 0.1 M citric acid, pH 4.0	1.6 M $(NH_4)_2SO_4$ 0.1M Bicine, pH 9.0	2.0M $(NH_4)_2SO_4$ 2% PEG 400 0.1M HEPES pH7.5
Resolution, (Å)	2.23 (2.28-2.23)	2.12 (2.14-2.12)	2.51 (2.58-2.51)	3.51 (3.86-3.51)
Space Group	R32	$P4_32_12$	R32	R32
a, (Å)	187.8	175.4	186.3	185.8
b, (Å)	187.8	175.4	186.3	185.8
c, (Å)	158.4	206.2	159.7	159.8
Molecules/asymmetric unit	3	12	3	3
Unique Reflections	52150	180816	35939	16306
Multiplicity	4.5 (3.9)	14.7 (11.7)	3.6 (3.3)	7.4 (7.3)
Average I/σ(I)	13.2 (1.0)	16.0 (2.4)	14.3 (0.7)	4.2 (1.0)
Rsym,(%)	8.7(155.7)	21.5(104.7)	10.8 (138.8)	52.2 (199.6)
Rpim (%)	4.1(77.5)	5.8(31.8)	6.6 (87.8)	18.2 (72.6)
$CC_{1/2}$ (%)	46.5	79.4	34.9	41.7
Completeness, (%)	100 (100)	100 (99.7)	98.9 (99.5)	100 (100)
Refinement				
Rwork, (%)	18.8	15.7	17.7	21.1
Rfree, (%)	21.33	19.4	21.4	30.8
RMSD from ideal				
Bond lengths, (Å)	0.008	0.008	0.008	0.006
Bond Angles, (°)	0.853	1.02	1.14	0.796
Numbers of atoms:				
Protein (mean ADP , $Å^2$)	4340 (46.4)	16953 (25.2)	4256 (69.6)	4287 (73.2)
Water (mean ADP, $Å^2$)	252	1596 (35.9)	84 (60.9)	-
Ligand (mean ADP, $Å^2$)	-	510 (24.2)	76 (76)	58 (58.94)
Ramachandran				
Favored/ Allowed/Outliers(%)	98.6/1.4/0.0	99.3/0.7/0.0	98.7/1.2/0.0	94.5/5.0/0.5

Table S2. Docking models and scores.

	LONP1					β5 Subunit (Chymotrypsin-like)					β1 Subunit (Caspase-like)					β2 Subunit (Trypsin-like)				
	Crystal Structures		Enlarged pocket models		Final Predicted Mode & Model	Crystal Structures		Enlarged pocket models		Final Predicted Mode & Model	Crystal Structures		Enlarged pocket models		Final Predicted Mode & Model	Crystal Structures		Enlarged pocket models		Final Predicted Mode & Model
	Apo* (6WYS)	Bortez-bound (6X27)	Apo MD sim*	Holo MD sim		Apo* (5LE5)	Bortez-bound (5LF3)	Apo MD sim*	Holo MD sim		Apo (5LE5)	Bortez-bound (5LF3)	Apo MD sim*	Holo MD sim		Apo (5LE5)	Bortez-bound (5LF3)	Apo MD sim*	Holo MD sim	
Pocket Volume (Å^3):	212.7	223.7	382.6	331.5		149.2	333.8	162.0	309.9		176.5	222.4	271.1	278.7		238.9	267.2	316.7	147.1	
Compound																				
bortezomib	-18.9	-26.3	-23.3	-27.2	**-27.2**	-25.8	-24.5	-6.5	-11.6	**-24.5**	-22.6	-29.4	-18.7	-41.4	**-41.4**	-18.3	-24.2	3.7	-18.2	**-24.2**
RMSD	7.0	1.0	11.7	1.0	**1.0**	3.6	1.7	3.0	1.9	**1.7**	0.7	2.6	9.0	0.8	**0.8**	2.3	2.4	7.7	1.2	**2.4**
5a	-14.7	-22.7	-11.9	-31.7	**-31.7**	-26.8	-19.1	-8.2	-10.3	**-19.1**	-20.3	-25.7	-13.2	-28.6	**-28.6**	-5.6	-10.3	-1.7	-22.4	**-22.4**
5b	-24.2	-18.9	-15.6	-26.7	**-18.9**	-20.8	-14.8	-5.7	-12.2		-22.2	-13.2	-11.3	-26.5	**-26.5**	-10.4	-14.5	-1.1	-17.7	**-17.7**
9a	-13.6	-20.8	-13.0	-22.5	**-20.8**	-11.5	-20.1	-7.2	-10.3	**-20.1**	-26.7	-22.6	-4.2	-35.5	**-35.5**	-10.2	-2.4	3.4	-13.7	**-13.7**
RMSD	5.8	1.1	6.8	6.6	**1.1**															
9b	-17.7	-20.9	-17.8	-24.8	**-24.8**	-16.8	-20.4	-14.5	-7.5	**-20.4**	-25.9	-21.6	-3.2	-29.2	**-29.2**	-5.3	-13.9	-1.6	-19.0	**-19.0**
9c	-15.5	-21.9	-14.4	-23.8	**-23.8**	-10.8	-18.8	-3.4	-4.7	**-18.8**	-15.8	-25.4	-3.5	-33.3	**-33.3**	-3.4	-13.1	2.2	-16.7	**-13.1**
9d	-18.1	-22.7	-18.9	-18.5	**-22.7**	-12.0	-18.1	-9.8	-0.5	**-18.1**	-20.7	-22.3	-4.6	-30.0	**-30.0**	-4.7	-3.1	3.6	-11.8	**-11.8**
9e	-5.4	-20.2	-16.5	-25.4	**-25.4**	-7.7	-15.9	-9.2	-12.5	**-15.9**	-18.6	-25.1	-5.9	-31.4	**-31.4**	-7.9	-11.2	9.2	-12.2	**-12.2**
9f	-15.8	-23.0	-10.7	-24.9	**-24.9**	-19.0	-18.3	-5.8	-6.9	**-18.3**	-20.1	-18.1	-4.2	-21.5	**-21.5**	-6.6	0.1	9.5	-13.2	
9g	-18.0	-17.1	-11.5	-16.0	**-17.1**	-4.2	-14.1	-0.4	-3.4	**-14.1**	-12.0	-21.6	-4.0	-17.8	**-17.8**	0.3	-7.5	7.8	-10.7	**-10.7**
9h	-9.2	-24.7	-12.5	-19.4	**-24.7**	-8.7	-20.2	-10.8	-15.5	**-20.2**	-20.6	-26.8	-4.1	-31.9	**-31.9**	-2.9	-14.9	2.9	-12.3	**-14.9**
9i	-18.8	-23.1	-18.2	-17.1	**-23.1**	-17.3	-11.6	-10.7	-7.9	**-11.6**	-15.6	-12.6	-1.4	-29.1	**-29.1**	-5.3	-11.9	5.5	-8.7	**-11.9**
11a	-22.0	-15.6	-14.1	-28.1	**-28.1**	-16.4	-1.3	3.6	-4.4		-7.0	-11.1	-7.0	-25.7	**-25.7**	-6.9	-7.7	5.5	-3.1	
11b	-14.1	-8.8	-21.4	-25.0		-13.3	-12.2	-3.3	-11.3	**-11.3**	-20.1	-2.0	-10.3	-22.3	**-22.3**	-6.5	-5.6	2.4	-1.3	
11c	-10.0	-14.2	-22.7	-23.4	**-22.7**	-9.0	-8.8	-9.2	-10.3	**-10.3**	-4.7	-4.1	-10.4	-28.3	**-28.3**	-15.1	-10.2	10.7	-10.3	**-15.1**
11d	-18.9	-12.6	-25.2	-29.3	**-29.3**	-16.1	-0.8	-2.1	-22.2	**-22.2**	-14.1	-6.5	-7.6	-33.2	**-33.2**	-3.5	-8.1	0.8	-11.5	
11e	-16.5	-17.9	-14.7	-25.1	**-14.7**	-20.7	-18.6	-8.6	-18.7	**-18.7**	-11.6	-10.1	-11.2	-25.6	**-25.6**	-9.8	-7.0	-9.8	-11.6	
11f	-18.3	-15.9	-19.5	-24.5	**-15.9**	-15.2	-14.7	-12.0	-14.4	**-14.4**	-8.5	-14.3	-11.7	-33.0	**-33.0**	-8.5	-14.4	3.7	-18.3	**-18.3**
12a	-6.8	-15.7	-13.0	-21.2	**-21.2**	-13.5	-11.4	-5.8	-4.1		-12.9	-6.2	-5.8	-23.3	**-23.3**	-2.8	-7.6	10.7	-9.2	
12d	-20.3	-10.4	-22.8	-26.5	**-26.5**	-16.4	-8.4	-0.5	-11.7	**-11.7**	-17.1	-8.7	-6.8	-26.9	**-26.9**	-3.5	-0.5	-1.6	-8.9	
12e	-22.3	-13.8	-22.0	-32.1	**-32.1**	-8.5	-11.8	-0.5	-11.0		-18.5	-5.9	-3.3	-25.8	**-25.8**	-4.8	-4.1	3.1	-1.3	
12f	-19.3	-17.1	-15.0	-21.0	**-21.0**	-13.1	-20.3	-12.2	-13.6	**-13.6**	-14.8	-14.2	-5.0	-30.9	**-30.9**	-17.1	-11.0	0.5	-13.5	**-17.1**
12g	-18.3	-17.6	-15.2	-22.7	**-22.7**	-11.0	-9.7	-1.9	-11.0	**-11.0**	-15.0	-20.9	-4.1	-29.9	**-29.9**	-11.5	-12.3	2.7	-16.6	**-16.6**
14	-15.3	-17.6	-23.0	-26.7	**-23.0**	-19.8	-10.5	-8.0	-16.5	**-16.5**	-8.2	-14.6	-7.4	-28.4	**-28.4**	-10.8	-5.4	0.5	-15.4	**-10.8**
Average score:	-16.3	-18.3	-17.2	-24.3	**-23.6**	-14.8	-14.3	-6.4	-10.5	**-16.5**	-16.4	-16.0	-7.0	-28.7	-28.7	-7.6	-9.2	3.0	-12.4	**-15.6**

* model eliminated due to poor posing/scoring of bortezomib test compound

Bortezomib-like pose | 9a-like pose | Alternative Pose

Table S3. Molecular formula strings.

compound	SMILE	LONP1 IC50 (uM)	20S proteasome IC50 (uM)
bortezomib	CC(C)C[C@H](NC(=O)[C@H](CC1=CC=CC=C1)NC(=O)C2=CN=CC=N2)B(O)O	0.183 ± 0.150	0.097 ± 0.072
5a	CC(C)C[C@@H](NC(=O)[C@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1)B(O)O	2.111 ± 0.129	1.524 ± 0.268
5b	CC(C)C[C@@H](NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C2=NC=CN=C2)B(O)O	2.980 ± 0.134	1.701 ± 0.235
9a	CC(C)C[C@H](NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1)B(O)O	0.253 ± 0.161	>10
9b	CC(C)C[C@H](NC(=O)[C@@H](C)NC(=O)C1=CN=CC=N1)B(O)O	0.433 ± 0.161	2.676 ± 0.905
9c	CC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CC(C)C)B(O)O	0.408 ± 0.212	>10
9d	CCC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CC(C)C)B(O)O	0.187 ± 0.092	>10
9e	CCCC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CC(C)C)B(O)O	0.093 ± 0.036	>10
9f	CC(C)C[C@H](NC(=O)[C@H](NC(=O)C1=CN=CC=N1)C(C)C)B(O)O	0.546 ± 0.258	3.971 ± 4.608
9g	CC(C)C[C@H](NC(=O)[C@H](NC(=O)C1=CN=CC=N1)C(C)(C)C)B(O)O	2.850 ± 0.632	>10
9h	CC(C)C[C@H](NC(=O)[C@@H](CC1CCCCC1)NC(=O)C1=CN=CC=N1)B(O)O	0.137 ± 0.077	0.843 ± 0.182
9i	CC(C)C[C@H](NC(=O)C(C)(C)NC(=O)C1=CN=CC=N1)B(O)O	6.199 ± 0.032	n.d.
11a	OB(O)[C@H](CC1CCCCC1)NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1	0.092 ± 0.015	0.541 ± 0.582
11b	CC(C)CCC[C@H](NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1)B(O)O	0.065 ± 0.013	1.036 ± 0.490
11c	OB(O)[C@H](CCCCCBr)NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1	0.077 ± 0.020	1.943 ± 0.938
11d	OB(O)[C@H](CCCC1=CC=CC=C1)NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1	0.018 ± 0.004	0.259 ± 0.194
11e	OB(O)[C@H](CCC1=CC=CC=C1)NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1	0.034 ± 0.015	0.262 ± 0.051
11f	OB(O)[C@H](CC1=CC=CC=C1)NC(=O)[C@@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1	0.109 ± 0.032	0.679 ± 0.227
12a	CCC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CC1CCCCC1)B(O)O	0.136 ± 0.134	>10
12d	CCC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CCCC1=CC=CC=C1)B(O)O	0.017 ± 0.012	>10
12e	CCC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CCC1=CC=CC=C1)B(O)O	0.092 ± 0.054	>10
12f	CCC[C@@H](NC(=O)C1=CN=CC=N1)C(=O)N[C@@H](CC1=CC=CC=C1)B(O)O	0.556 ± 0.354	>10
12g	CCCC[C@H](NC(=O)[C@@H](CCC)NC(=O)C1=CN=CC=N1)B(O)O	0.038 ± 0.010	>10
14	CCC[C@@H](NC(=O)C1=C(C)N=C(C)O1)C(=O)N[C@@H](CCCC1=CC=CC=C1)B(O)O	0.059 ± 0.046	>10

Table S4. Nanosyn protease panel of **14**. Percent inhibition measured at 1 μM compound concentration.

Protease	Mean %-inh (n=2)	Protease	Mean %-inh (n=2)	Protease	Mean %-inh (n=2)
ACE	-1	MMP2	-13	DPP8	-6
ACE2	-2	MMP3	1	DPP9	12
ADAM10	1	MMP7	-2	Factor VII	1
BACE-1	-6	MMP8	-3	Factor-Xa	-2
Calpain-1	-10	MMP9	11	FAP	9
Caspase-1	17	Neprilysin	2	Furin	4
Caspase-2	1	Plasma-Kallikrein	3	Granzyme-A	-1
Caspase-3	1	Plasmin	-3	Granzyme-B	-2
Caspase-4	1	Prolyl Oligopeptidase	2	Granzyme-K	-1
Caspase-5	-1	PSMB10	-6	HTRA2	1
Caspase-6	2	PSMB5	9	IDE	-8
Caspase-7	7	PSMB6	-7	Kallikrein11	6
Caspase-8	1	PSMB7	-31	Kallikrein13	3
Caspase-9	9	PSMB8	22	Kallikrein5	-6
Cathepsin-B	-23	PSMB9	4	Kallikrein7	9
Cathepsin-D	11	Spinesin	14	Matriptase	-4
Cathepsin-K	-14	TACE	5	MMP1	7
Cathepsin-L	-4	Thrombin	2	MMP12	0
Cathepsin-S	3	tPA	5	MMP13	21
DPP3	6	uPA	0	MMP14	5
DPP4	-7	ACE	0	MMP2	6
DPP8	-5	ACE2	-4	MMP3	0
DPP9	3	ADAM10	2	MMP7	-2
Factor VII	4	BACE-1	-9	MMP8	-13
Factor-Xa	1	Calpain-1	2	MMP9	28
FAP	0	Caspase-1	7	Neprilysin	3
Furin	-1	Caspase-2	3	Plasma-Kallikrein	4
Granzyme-A	-4	Caspase-3	1	Plasmin	6
Granzyme-B	3	Caspase-4	-1	Prolyl Oligopeptidase	-1
Granzyme-K	-10	Caspase-5	-1	PSMB10	-7
HTRA2	6	Caspase-6	6	PSMB5	17
IDE	-3	Caspase-7	6	PSMB6	-2
Kallikrein11	22	Caspase-8	2	PSMB7	-38
Kallikrein13	3	Caspase-9	9	PSMB8	52
Kallikrein5	-5	Cathepsin-B	-12	PSMB9	4
Kallikrein7	3	Cathepsin-D	4	Spinesin	1
Matriptase	-3	Cathepsin-K	-7	TACE	4
MMP1	4	Cathepsin-L	0	Thrombin	1
MMP12	11	Cathepsin-S	2	tPA	9
MMP13	18	DPP3	6	uPA	-1
MMP14	3	DPP4	-9		

SUPPORTING FIGURES

Figure S1.

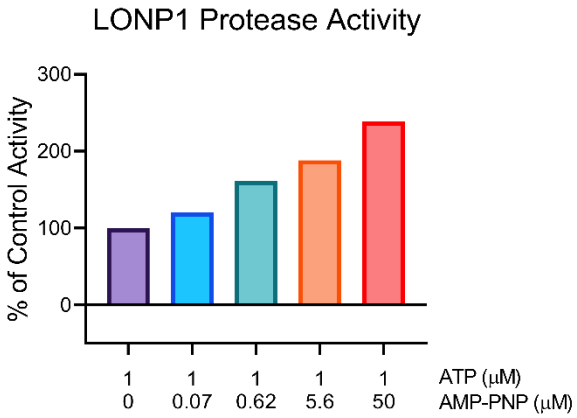


Figure S1. Biochemical characterization of human LONP1 depicting the effect of AMP-PNP on LONP1 protease activity. Assay containing 1 μM ATP was defined as 100% control activity.

Figure S2.

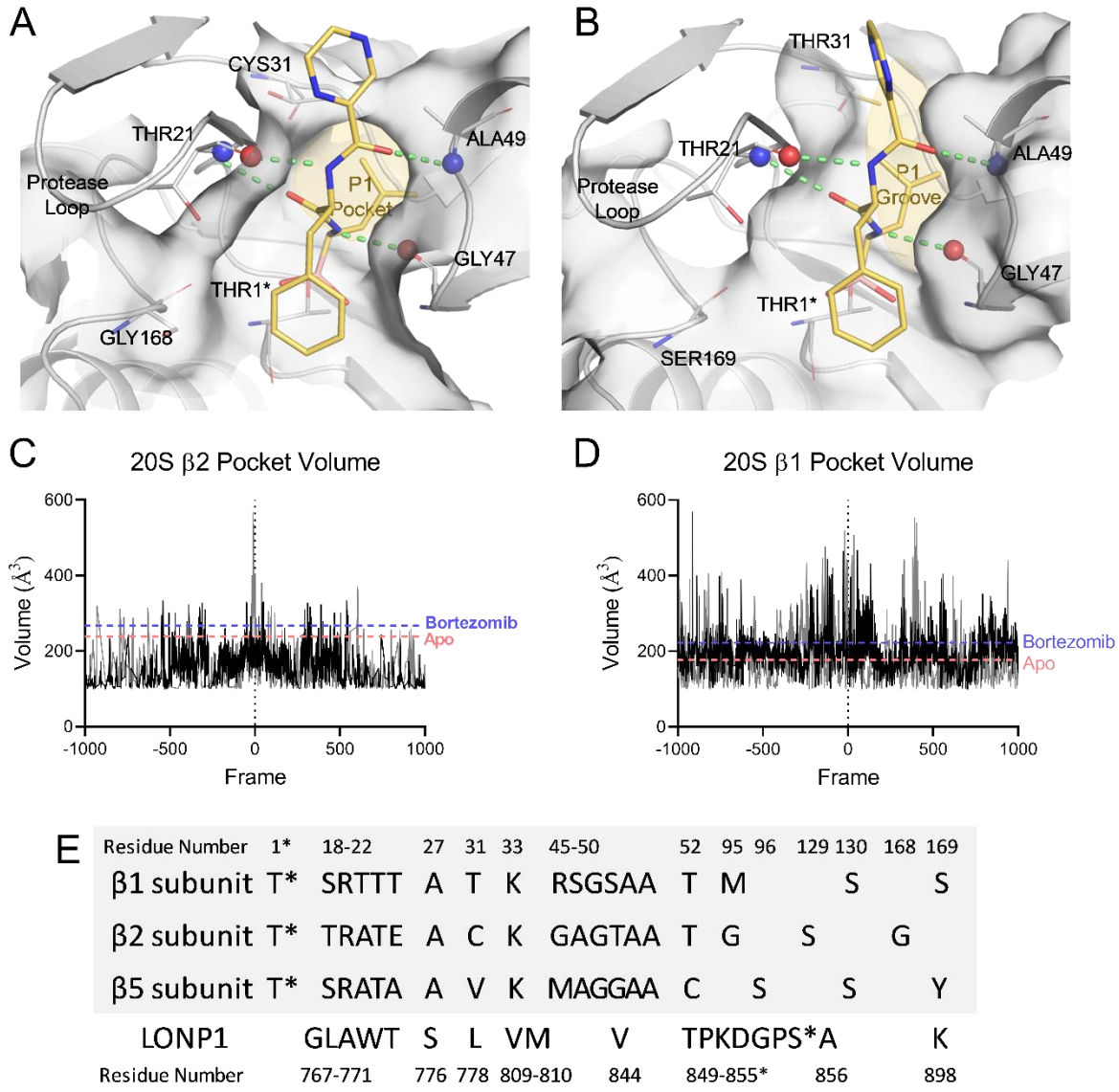


Figure S2. Active sites of β2 caspase-like (**A**) and β1 trypsin-like (**B**) 20S proteasome subunits. MD simulation of 20S proteasome protease domains were carried out in duplicate. The volume of the binding site was calculated over the duration of the simulation for the β2 caspase-like (**C**) and β1 trypsin-like (**D**) 20S proteasome subunits. The volume of the binding site in the apo and bortezomib-bound crystal structures is also plotted. (**E**) Structural alignment of binding site

residues in LONP1 and 20S proteasome subunits. The catalytic serine and threonine of LONP1 and 20S proteasome, respectively, are annotated with a *.

Figure S3.

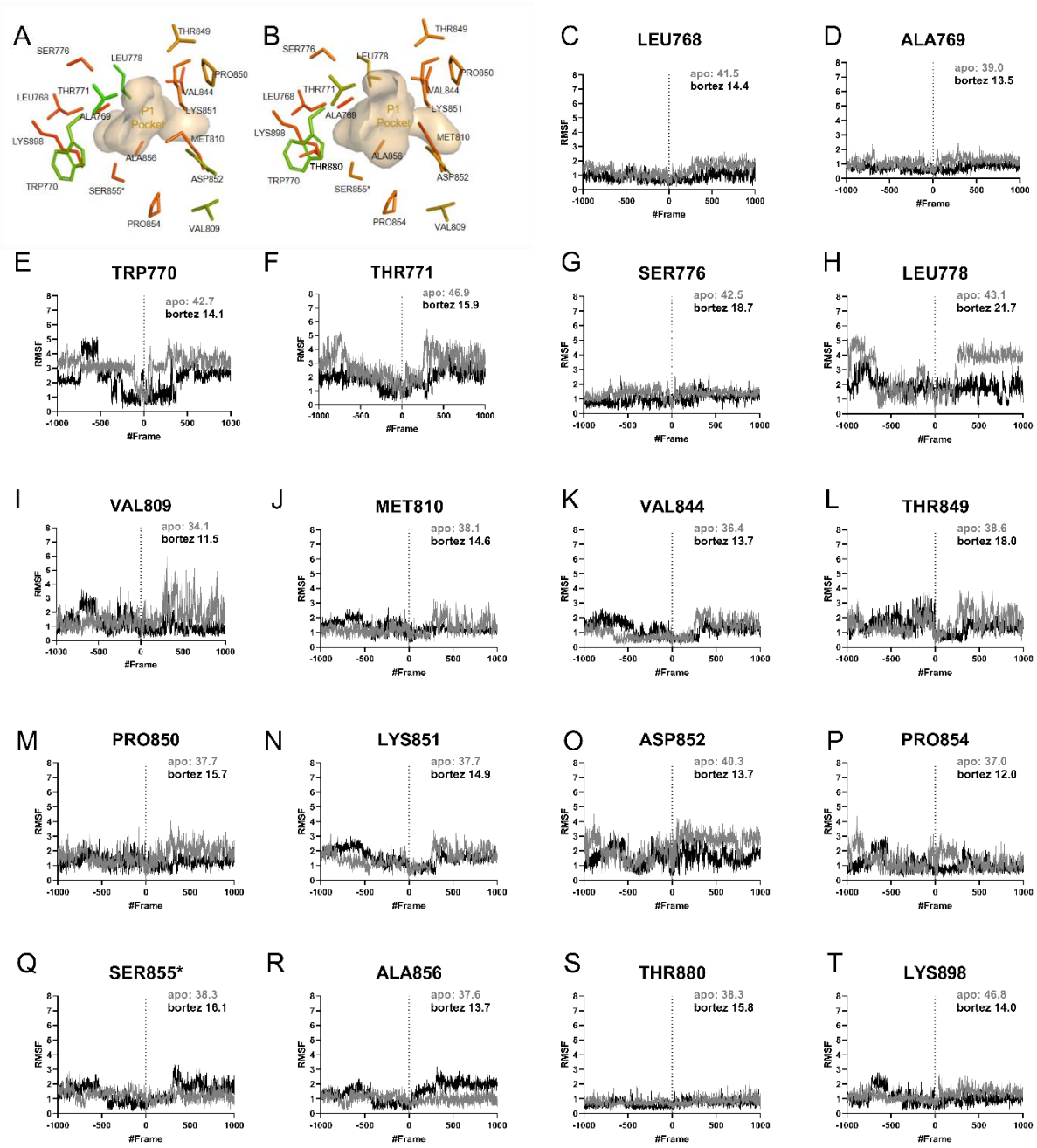


Figure S3. Binding site residue dynamics from 200 ns of MD simulation of the apo and bortezomib-bound LONP1 protease domains. The binding site architecture of the apo (**A**) and bortezomib-bound (**B**) structures are shown for reference. Residues are colored according to

flexibility; those in green have a broader RMSF range than those in red. The P1 portion of the pocket is highlighted. The RMSF of individual residues over the duration of each MD simulation is shown in panels **C-T**. The apo simulation data is in grey and the bortezomib simulations are in black. Glycine residues have been excluded. Residue average b-factors are also included in the upper right hand corner of each plot. The catalytic serine is annotated with a *.

Figure S4.

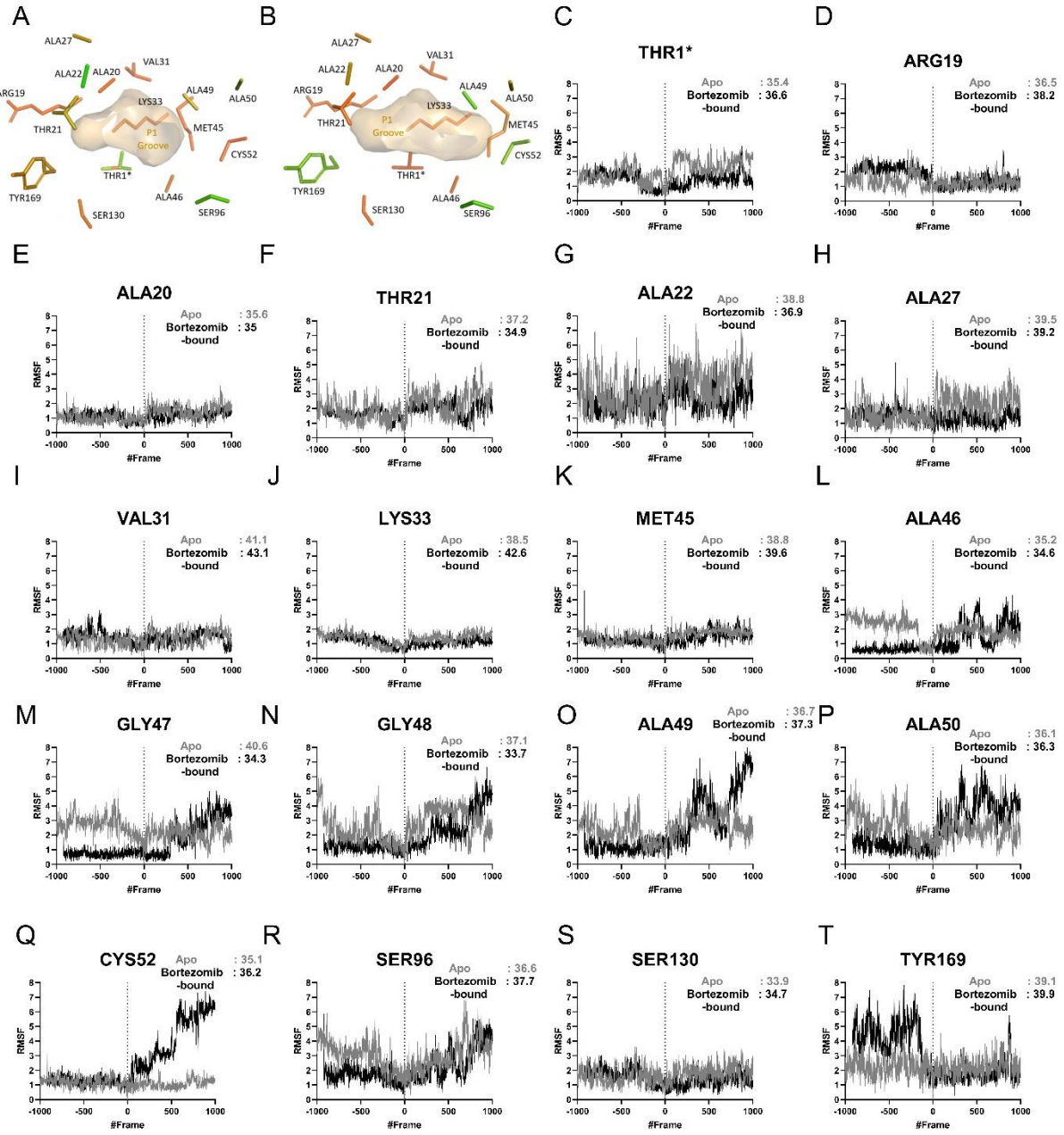


Figure S4. Binding site dynamics of apo and bortezomib-bound human 20S proteasome β5 protease domain from a 200 ns of MD simulation. The binding site architecture of the apo (**A**) and bortezomib-bound (**B**) structures are shown for reference. Residues are colored according to flexibility; those in green have a broader RMSF range than those in red. The P1 portion of the

pocket is highlighted. The RMSF of individual residues over the duration of each MD simulation is shown in panels **C-T**. The apo simulation data is in grey and the bortezomib simulations are in black. Glycine residues have been excluded. Residue average b-factors are also included in the upper right hand corner of each plot. The catalytic threonine is annotated with a *.

Figure S5.

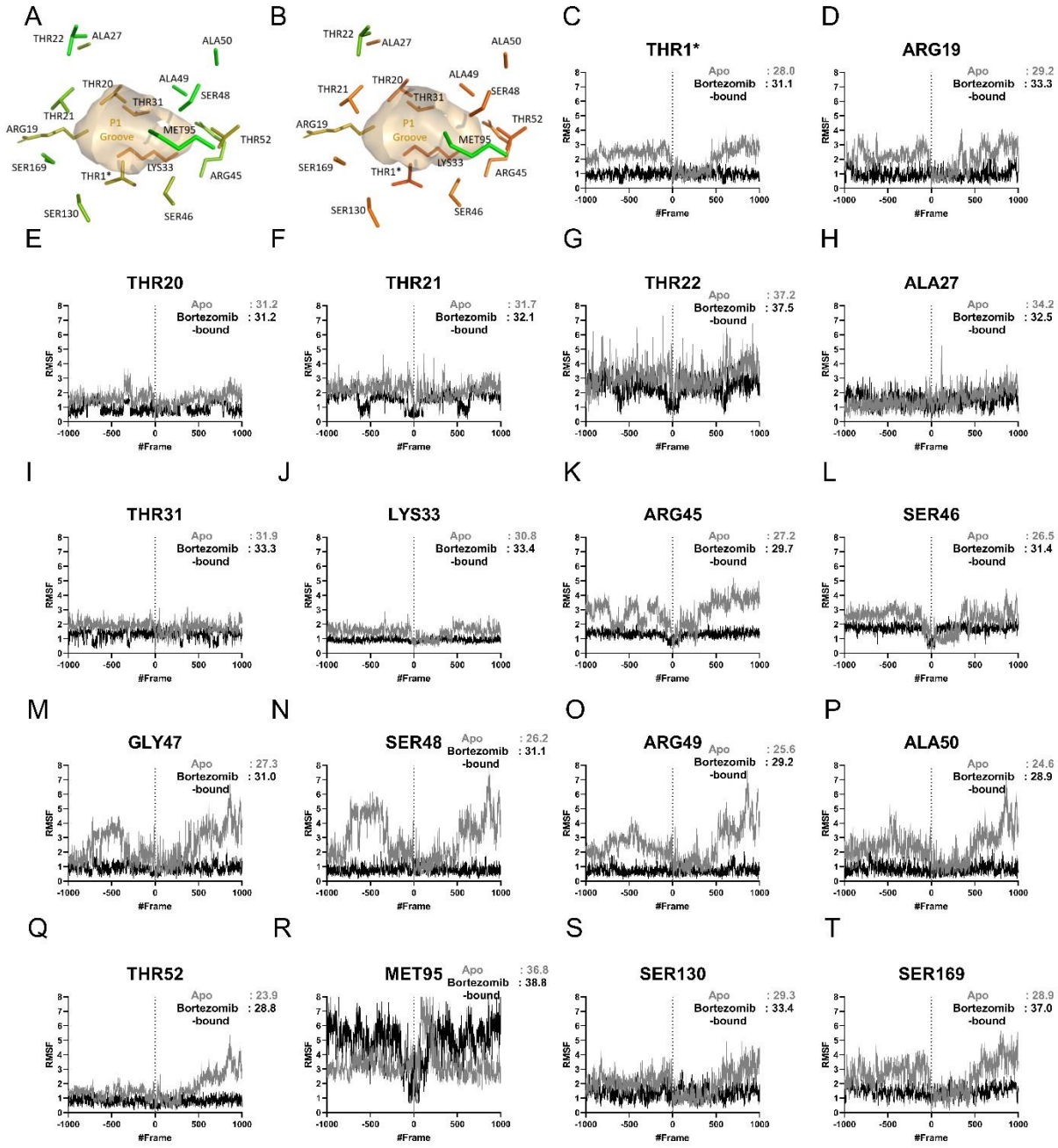


Figure S5. Binding site dynamics of apo and bortezomib-bound human 20S proteasome β1 protease domain from a 200 ns of MD simulation. The binding site architecture of the apo (**A**) and bortezomib-bound (**B**) structures are shown for reference. Residues are colored according to flexibility; those in green have a broader RMSF range than those in red. The P1 portion of the

pocket is highlighted. The RMSF of individual residues over the duration of each MD simulation is shown in panels **C-T**. The apo simulation data is in grey and the bortezomib simulations are in black. Glycine residues have been excluded. Residue average b-factors are also included in the upper right hand corner of each plot. The catalytic threonine is annotated with a *.

Figure S6.

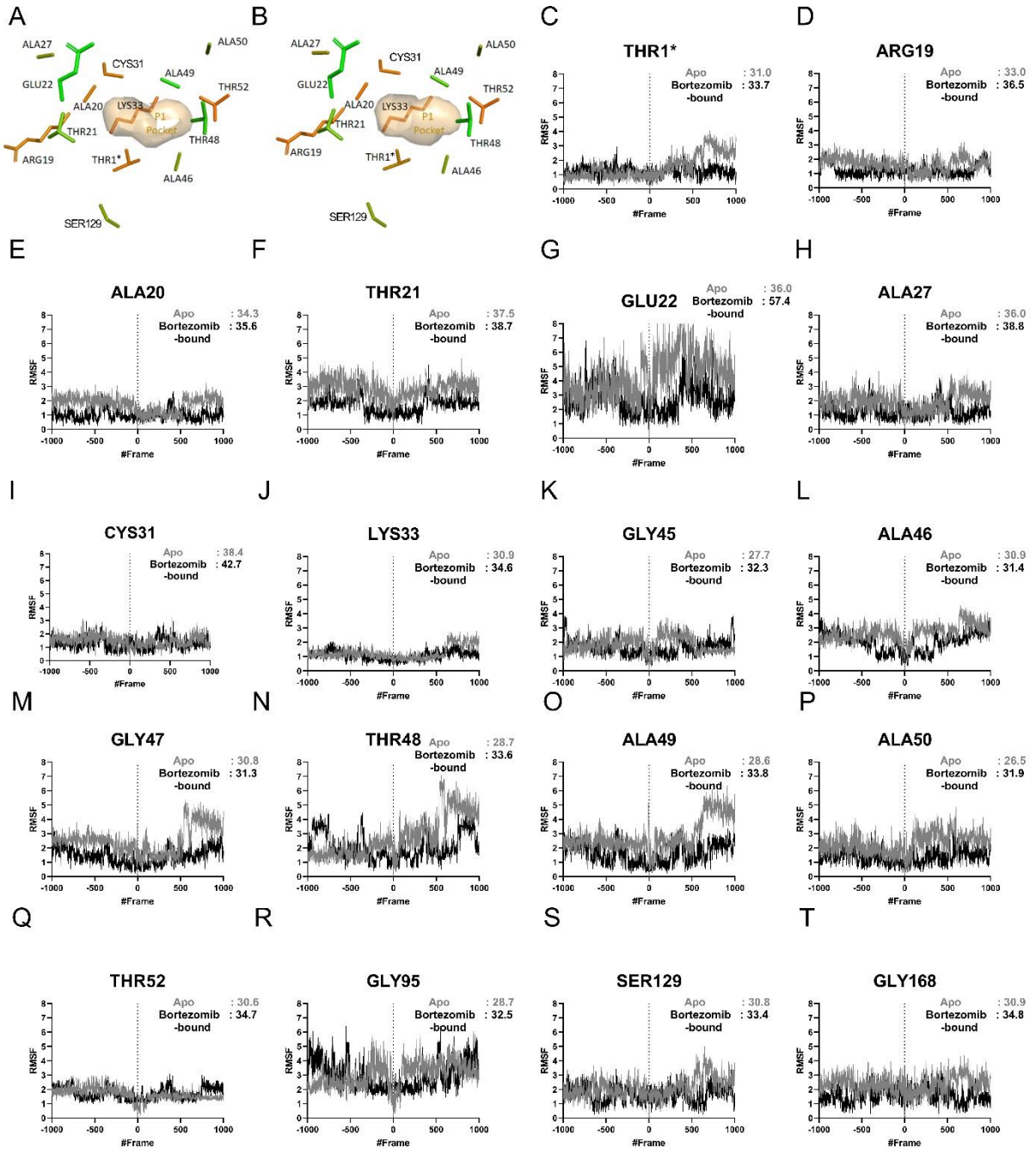
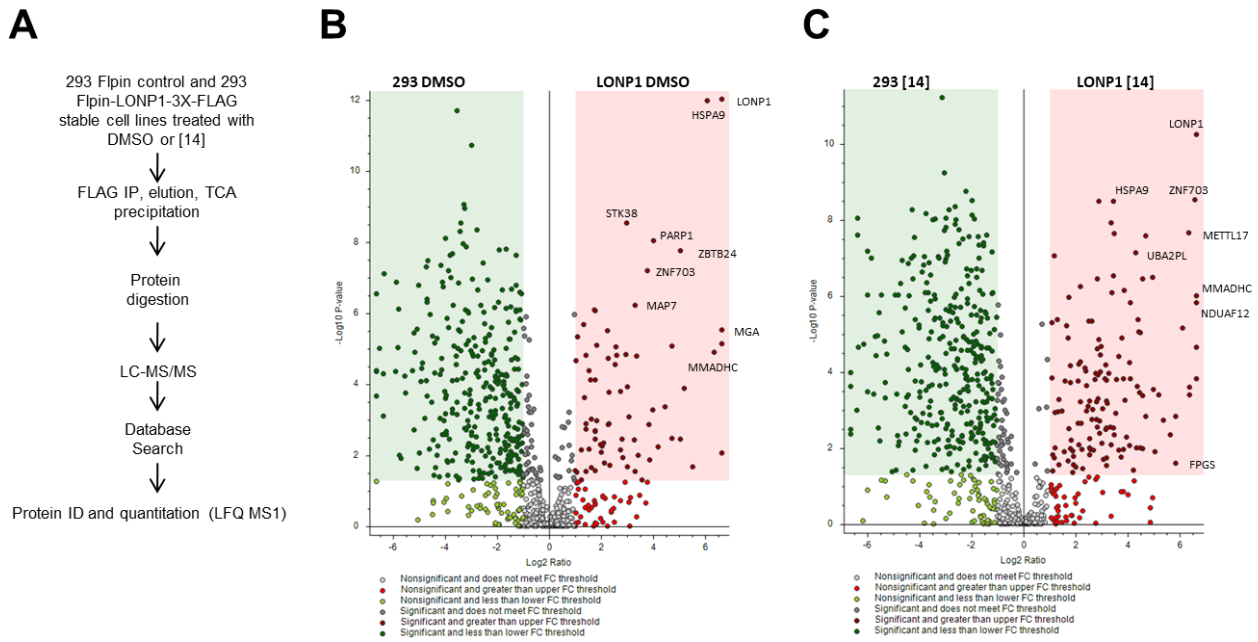


Figure S6. Binding site dynamics of apo and bortezomib-bound human 20S proteasome β2 protease domain from a 200 ns of MD simulation. The binding site architecture of the apo (**A**) and bortezomib-bound (**B**) structures are shown for reference. Residues are colored according to

flexibility; those in green have a broader RMSF range than those in red. The P1 portion of the pocket is highlighted. The RMSF of individual residues over the duration of each MD simulation is shown in panels **C-T**. The apo simulation data is in grey and the bortezomib simulations are in black. Glycine residues have been excluded. Residue average b-factors are also included in the upper right hand corner of each plot. The catalytic threonine is annotated with a *.

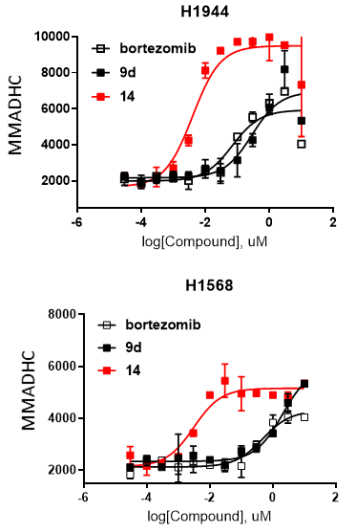
Figure S7.



D

Gene Symbol	# Peptides	# PSMs	# Unique Peptides	Abundance Ratio (log2): LONP DMSO/ 293 DMSO	Abundance Ratio (log2): LONP [14] / 293 [14]
LONP1	107	4935	106	6.64	6.64
HSPA9	58	1185	57	6.09	3.44
XRN2	29	100	29	1.36	3.44
CLUH	21	79	21	2.6	2.4
ZNF703	20	104	18	3.78	6.6
CIZ1	18	73	18	4.45	3.15
UBAP2L	18	92	18	2.28	4.3
HADHA	16	41	16	2.58	3.86
FBXO11	16	77	16	1.79	3.36
ELAC2	15	50	15	1.77	3.65
MMADHC	13	134	13	6.34	6.64
TARS2	9	12	9	2.73	6.36
ADCK3	8	28	8	2.35	4.45
KARS	8	20	8	1.76	1.75
CRTC3	8	26	8	1.43	1.47
SDHB	7	16	7	3.99	3.2
TRAP1	7	13	6	1.07	4.38
NTPCR	6	23	6	4.74	4.47
CUL1	5	11	5	2.96	2.99
MRPS2	5	14	5	1.63	2.11
SAMHD1	5	8	5	1.24	2.09
MGA	4	6	4	6.64	6.64
FPGS	4	9	4	6.64	5.84
POLG	4	12	4	3.86	3.2
MARS	4	5	4	3.3	2.54
FARSA	4	17	4	2.55	2.36
SQSTM1	4	15	4	2.1	2.41
ERC2	4	9	4	1.78	2.78
MRPS18C	3	6	3	3.41	3.48
UBR4	3	9	3	3.36	2.88
UBA1	3	5	3	3.01	3.29
COQ5	3	5	3	2.03	1.12
NDUFAF3	3	8	3	1.77	2.11
ZBTB24	2	5	2	5.05	4.7
SF3B5	2	10	2	4.72	6.64
PARP1	2	2	2	4.01	3.49
PSMD13	2	4	2	3.07	2.29
MRPS26	2	6	2	3.01	3.02
PSMD6	2	5	2	2.85	3.32
MKRN2	2	9	2	2.74	3.79
METTL15	2	6	2	2.29	2.42

E

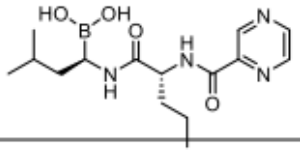


Compound	Cell Line	EC50 (nM)
bortezomib	Calu6	152
	H1568	475
	H1944	152.2
9d	Calu6	143.3
	H1568	2007
	H1944	672
14	Calu6	3.2
	H1568	3.1
	H1944	4.7

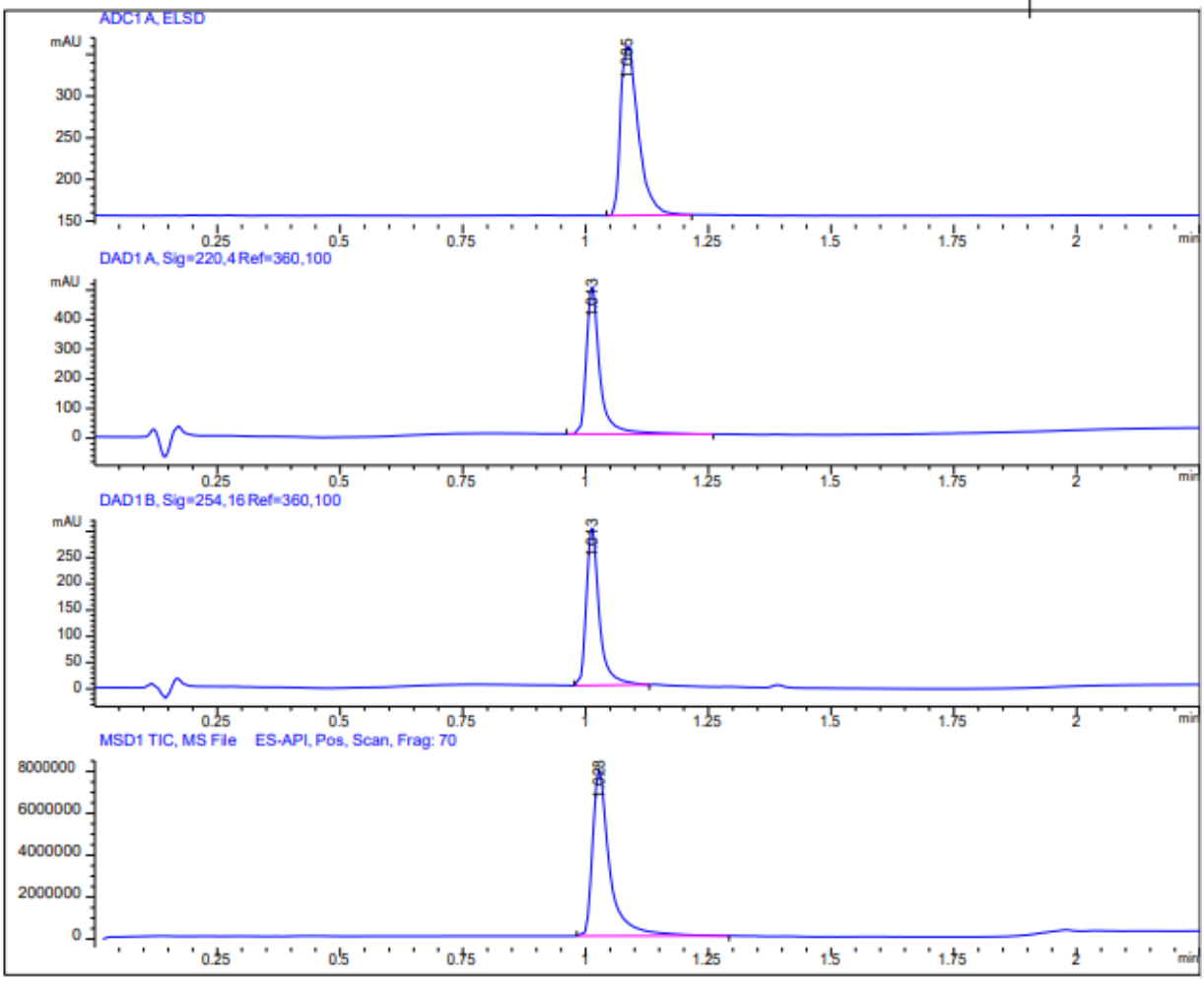
Figure S7. (**A**) Experiment design of affinity purification and mass spectrometry analysis to identify LONP1 binding partners. Volcano plots showing the proteins that were enriched more than 2-fold and met the p-value cutoff from DMSO-treated (**B**) and **14**-treated (**C**) HEK293 control and LONP1-3XFLAG stable cell lines. The pink rectangle shows the proteins enriched in LONP1 samples whereas the green rectangle shows the proteins enriched in control samples. (**D**) Table listing the top 25 proteins that were enriched from DMSO and **14**-treated samples. MMADHC, highlighted in red, was observed to be an abundant protein that was further enriched upon treatment with **14**. (**E**) MMADHC measured by HTRF assay in H1944 and H1568 cells following treatment with indicated compounds. A table summarizing the MMADHC HTRF data from Calu6, H1944, and H1568 is shown below.

Figure S8.

((*R*)-3-methyl-1-((*R*)-2-(pyrazine-2-carboxamido)pentanamido)butyl)boronic acid **9d**

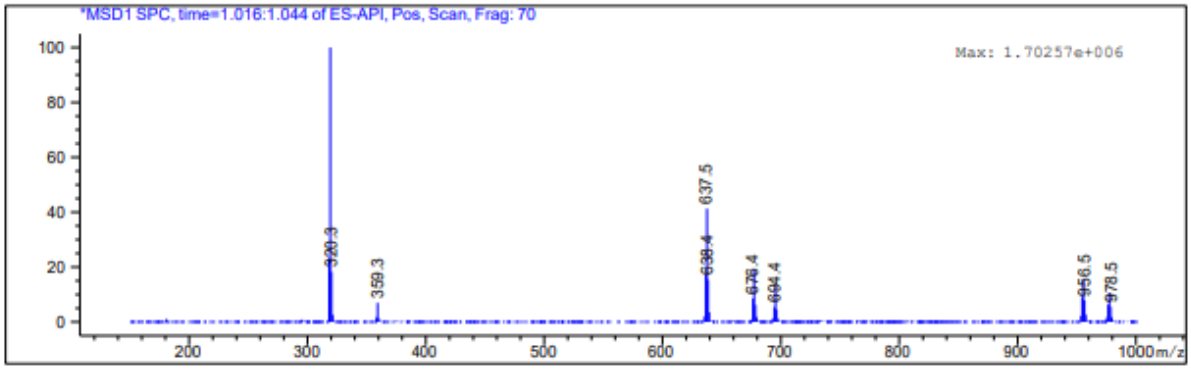


```
Acq. Instrument : Instrument 1                Location: Vial 8
Inj :1                                        Inj Volume: 3.000 µl
Acq. Method     : C:\CHEM32\1\METHODS\2MIN_10TO100B_150-1
Analysis Method : C:\CHEM32\1\METHODS\DEF_LC.M
```



MS Signal: MSD1 TIC, MS File, ES-API, Pos, Scan, Frag: 70
Spectra averaged over upper half of peaks.
Noise Cutoff: 1000 counts.
Reportable Ion Abundance: > 10%.

Retention Time (MS)	MS Area	Mol. Weight or Ion
1.028	19196332	977.50 I
		955.60 I
		695.40 I
		677.40 I
		638.40 I
		637.50 I
		636.45 I
		320.30 I
		319.30 I
		318.30 I



*** End of Report ***

Page 2 of 2

Figure S8. HPLC traces of **9d**.

Figure S9.

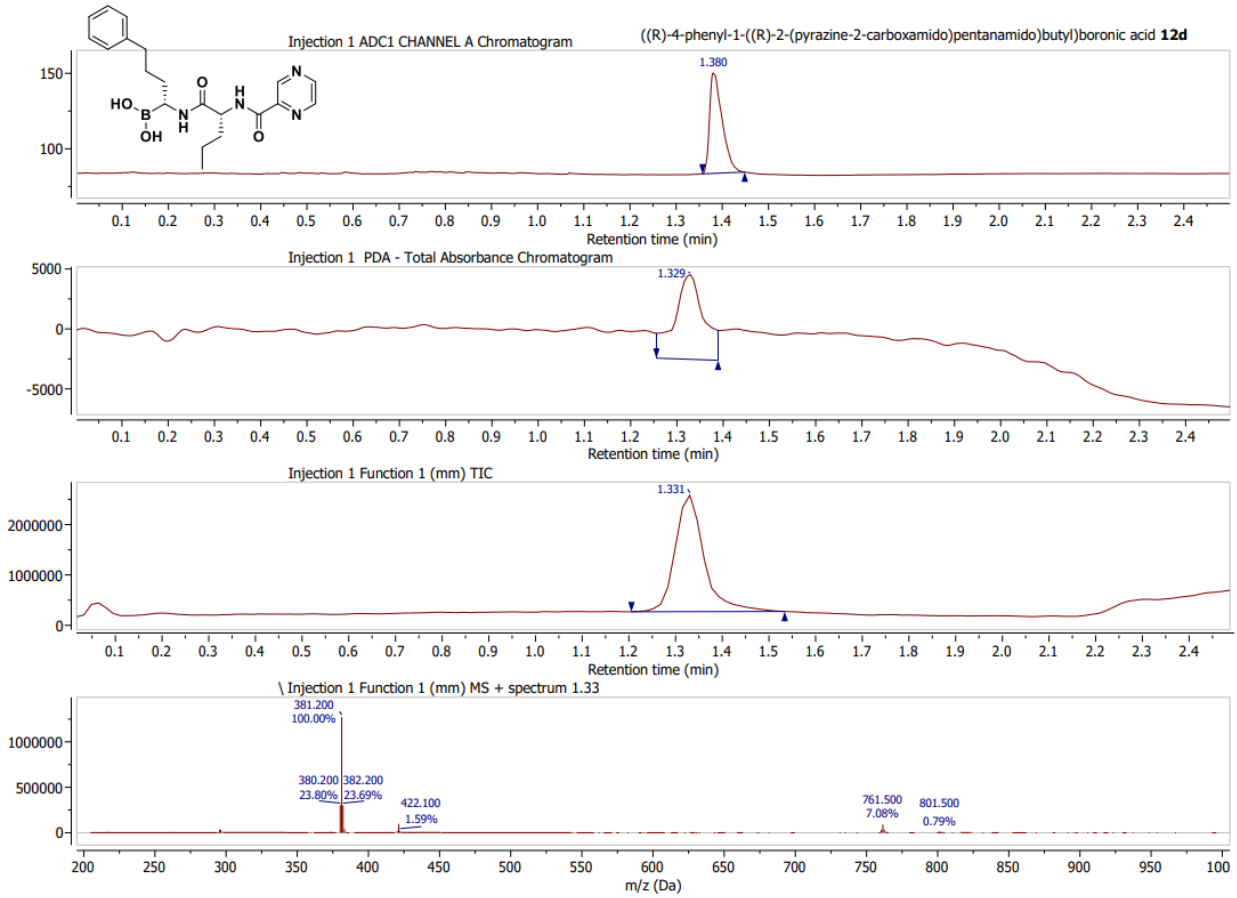
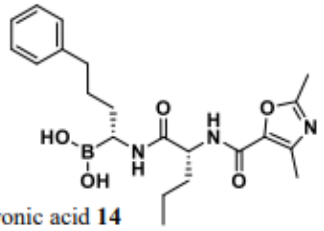


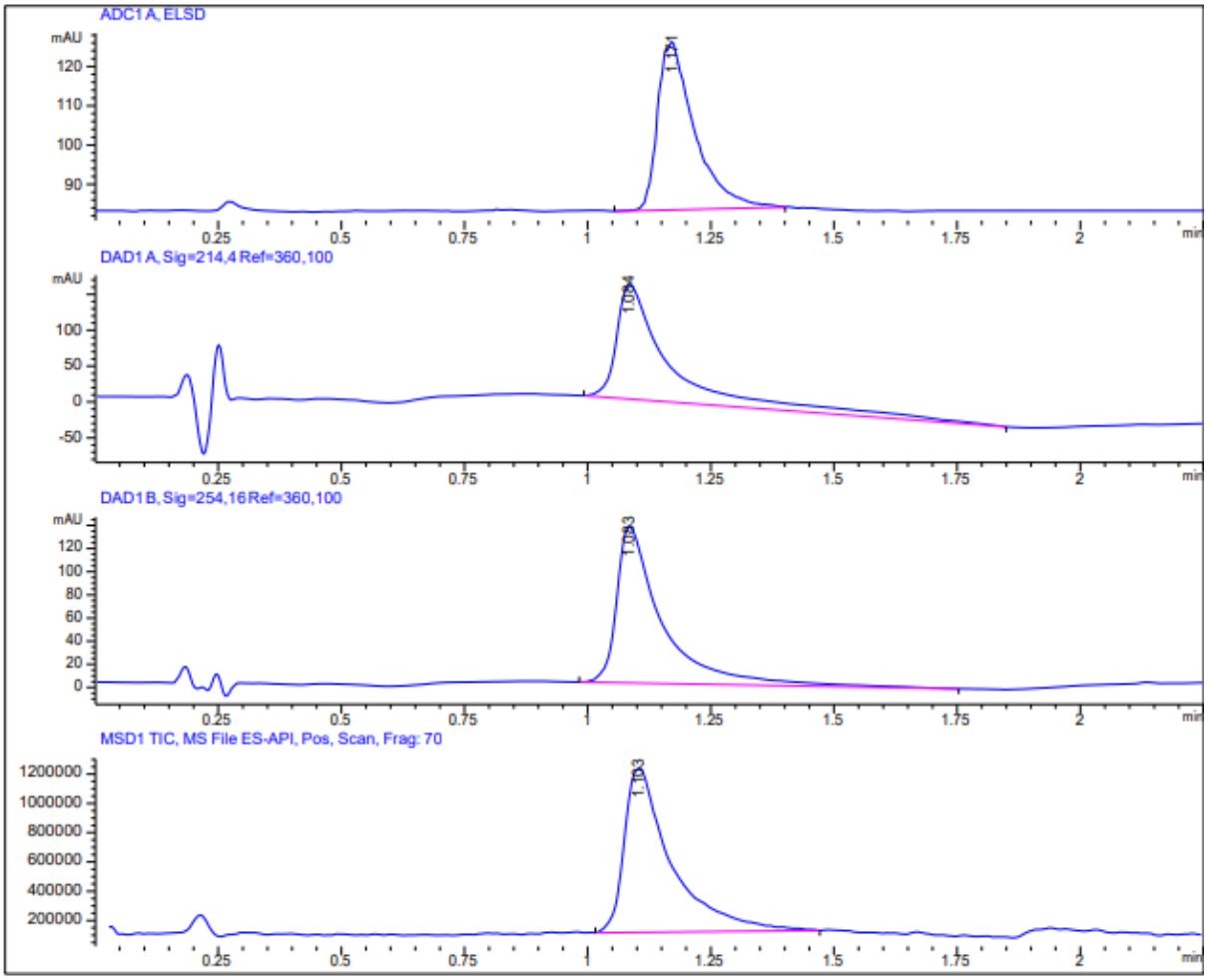
Figure S9. HPLC traces of **12d**.

Figure S10.



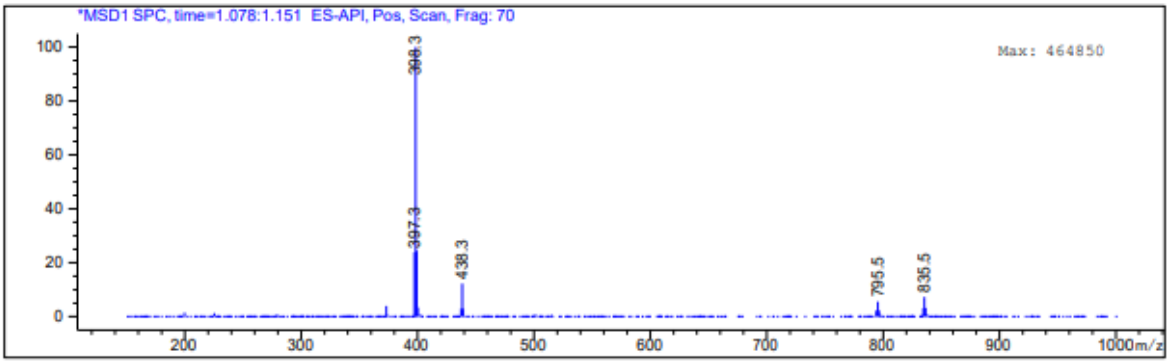
Analysis Method: C:\CHEM32\1\METHODS\DEF_LC.M

((*R*)-1-((*R*)-2-(2,4-dimethyloxazole-5-carboxamido)pentanamido)-4-phenylbutyl)boronic acid **14**



MS Signal: MSD1 TIC, MS File, ES-API, Pos, Scan, Frag: 70
Spectra averaged over upper half of peaks.
Noise Cutoff: 1000 counts.
Reportable Ion Abundance: > 10%.

Retention Time (MS)	MS Area	Mol. Weight or Ion
1.103	7454955	438.30 I
		399.30 I
		398.30 I
		397.30 I



*** End of Report ***

Figure S10. HPLC traces of **14**.