Supporting information for "Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19"

A. Acharya[1], R. Agarwal[2-4], M. Baker[5], J. Baudry[6], D. Bhowmik[7], S. Boehm[5], K. G. Byler[6], S.Y. Chen[9], L. Coates[8], C.J. Cooper[2,4], O. Demerdash[10], I. Daidone[11], J.D. Eblen[2,3], S. Ellingson[13], S. Forli[14], J. Glaser[15], J. C. Gumbart[1], J. Gunnels[16], O. Hernandez[5], S. Irle[7,17,18], D.W. Kneller[8], A. Kovalevsky[8], J. Larkin[19], T.J. Lawrence[10], S. LeGrand[19], S.-H. Liu[2,3], J.C. Mitchell[10], G. Park[9], J.M. Parks[2-4], A. Pavlova[1], L. Petridis[2,3], D. Poole[19], L. Pouchard[9], A. Ramanathan[20], D. Rogers[15], D. Santos-Martins[14], A. Scheinberg[21], A. Sedova[10], Y. Shen[2-4], J.C. Smith*[2,3], M.D. Smith[2,3], C. Soto[9], A. Tsaris[15], M. Thavappiragasam[10], A.F. Tillack[14], J.V. Vermaas[15], V.Q. Vuong[7,17,18], J.Yin[15], S. Yoo[9], M. Zahran[22], L. Zanetti-Polzi[23]

[1] School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2] UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, TN, 37830, USA
[3] The University of Tennessee, Knoxville. Department of Biochemistry & Cellular and Molecular Biology, 309 Ken and Blaire Mossman Bldg. 1311 Cumberland Avenue Knoxville, TN, 37996, USA
[4] Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN, 37996, USA
[5] Computer Science and Mathematics Division, Oak Ridge National Lab, Oak Ridge, TN 37830, USA
[6] The University of Alabama in Huntsville, Department of Biological Sciences. 301 Sparkman Drive, Huntsville, AL 35899, USA
[7]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[8]Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[9]Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973, USA
[10]Biosciences Division, Oak Ridge National Lab, Oak Ridge, TN 37830, USA
[11]Department of Physical and Chemical Sciences, University of L'Aquila, I-67010 L'Aquila, Italy
[13]University of Kentucky, Division of Biomedical Informatics, College of Medicine, UK Medical Center MN 150, Lexington KY, 40536, USA
[14]Scripps Research, La Jolla, CA, 92037, USA
[15]National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA
[16]HPC Engineering, Amazon Web Services, Seattle, WA 98121, USA
[17]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[18]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN 37996, USA
[19]NVIDIA Corporation, Santa Clara, CA 95051, USA
[20]Data Science and Learning Division, Argonne National Lab, Lemont, IL 60439, USA
[21]Jubilee Development, Cambridge MA 02139, USA
[22]Department of Biological Sciences, New York City College of Technology, The City University of New York (CUNY), Brooklyn, NY 11201, USA
[23] CNR Institute of Nanoscience, I-41125 Modena, Italy

[¥]Authors are listed in alphabetical order.

*Corresponding Author: Jeremy C. Smith, smithjc@ornl.gov

Table S1. Summary of simulated systems and the number of replicas per system.

PDB	**Name**	**Oligomerization State / Modeling Notes**	**# Replicas**
6W41	*spike protein*	*Receptor binding domain only ("Apo")*	*30*
6W41	*spike protein*	*Receptor binding domain in complex with ACE2*	*60*
6Y2E	*MPro (main protease)*	*Monomer with default CHARMM protonation (all HIS to HSD)*	*38*
6Y2E	*MPro (main protease)*	*Dimer with default CHARMM-GUI protonation (all HIS to HSD)*	*39*
6WQF	*MPro (main Protease)*	*Charged dimer. Protonation states: HSD41, HSP64, HSP80, HSP163, HSP164, HSE172, HSE246*	*40*
6WQF	*MPro (main protease)*	*Dimer. Protonation states: HSD41, HSD64, HSD80, HSE163, HSE164, HSE172, HSE246*	*40*
6WQF	*MPro (main protease)*	*Monomer. Protonation states: HSD41, HSD64, HSD80, HSE163, HSE164, HSE172, HSE246*	*38*
6WQF	*MPro (main protease)*	*Dimer. Protonation states: HSE41, HSD64, HSD80, HSE163, HSE164, HSE172, HSE246*	*40*
6WQF	*MPro (main protease)*	*Monomer. Protonation states: HSE41, HSD64, HSD80, HSE163, HSE164, HSE172, HSE246*	*38*

6W4H	*NSP10*	*Monomer, explicit $Zn(Cys)_4$ coordination*	*30*
6W4H	*NSP16*	*Monomer*	*27*
6W4H	*NSP10/NSP16, Methyltransferase complex*	*Heterodimer with explicit $Zn(Cys)_4$ coordination*	*40*
6VWW	*NSP15, Endoribonuclease*	*Monomer, His tags removed*	*36*
6VWW	*NSP15, Endoribonuclease*	*Hexamer, His-tags removed*	*57*
6M3M	*N protein, N-terminal RNA binding domain*	*Monomer, no RNA, no Zn*	*25*
6M3M	*N protein, N-terminal RNA binding domain*	*Tetramer, no RNA, no Zn*	*43*
6VYO	*N protein, N-terminal binding domain (Phosphoprotein)*	*Monomer, Zn^{2+} present in crystal, not present in simulation,*	*25*
6VYO	*N protein, N-terminal RNA binding domain (Phosphoprotein)*	*Tetramer, Zn^{2+} Complexed w/explicit bonds (patched)*	*39*
6W02	*NSP3, phosphatase domain*	*Asymmetric unit (dimer), tags not removed, Apo*	*35*
6W02	*NSP3, phosphatase domain*	*monomer*	*22*
6W4B	*NSP9*	*Monomer*	*23*
6W4B	*NSP9*	*Dimer*	*33*
6W9C	*PL protease*	*Charged Monomer, computationally refined structure. Zn^{2+} complexed w/explicit bonds (patched). PropKa based assignment Protonation States: HSP15, HSP45,*	*38*

		HSP48, HSP71, HSP87, HSD173, HSP253, HSP270, HSD273,	
6WRH	*PL protease*	*Neutral Monomer, Zn^{2+} complexed w/explicit bonds (patched). Neutral Protonation states (manual assignment): HSE17, HSE47, HSE50, HSE73, HSD89, HSD175, HSE255, HSE272, HSD275; C111S crystallization mutation removed and WT sequence used*	*38*

Table S2: Summary of the docking performed. The docking was performed on the top 10 conformations of the clusters calculated from 750ns TREMD simulations. The table provides information on the oligomeric state, the docking region, the residues used for clustering, the center of the docking box, and the docking box size.

Simulation	**Docking region**	**Clustering residues**	**Atom at the center of docking box**	**Docking box size**
NSP10 monomer (PDB Id: 6W4H)	Binding interface to NSP16	4253 to 4267, 4295 to 4304, 4309 to 4314, 4321 to 4324, 4337 to 4340, 4347 to 4349	OH TYR 4349	
			O LYS P4296	
NSP16 monomer (PDB Id: 6W4H)	Binding interface to NSP10	6834 to 6839, 6874 to 6878, 6880 to 6882, 6884 to 6886, 6888 to 6891, 6900 to 6907, 7042 to 7047	HE21 GLN 6885	
	SAM-binding site	6841, 6845, 6867, 6869 to 6872, 6877 to 6879, 6897 to 6899, 6911 to 6913, 6928 to 6931, 6947, 6968	CG ASP 6928	

Complex NSP10-NSP16 (PDB Id: 6W4H)	Complex interface	6832 to 6842, 6871 to 6886, 6888 to 6891, 6900 to 6907, 7041 to 7049, 7085 to 7095, 4253 to 4267, 4295 to 4304, 4309 to 4314, 4321 to 4324, 4337 to 4340, 4347 to 4349	OH TYR 4349	
			OH TYR 4349	
Tetramer N-terminal domain of nucleocapsid (N) protein **(PDB IDs: 6VYO)**	RNA binding sites	88, 92, 107	geometric center of ARG 88, ARG 92, and ARG 107	30 x 30 x 30 Å
Dimer Main Protease (All HSD) **(PDB Id: 6Y2E)**	dimer interface		-23.53, -3.61, -10.42	30 x 29 x 28 Å
Tetramer Nucleocapsid (N) Protein Complex **(PDB Id: 6VYO)**	Three binding sites predicted by FTMap		-23.53, -3.61, -10.42	30 x 29 x 28 Å
			-23.53, -3.61, -10.42	30 x 29 x 28 Å
			-23.53, -3.61, -10.42	30 x 29 x 28 Å
Nucleocapsid Monomer **(PDB Id: 6KL5)**			34.42, 28.40, 29.28	28 x 28 x 28 Å

MPro Monomer (three different protonation variants) (PDB Ids: 6WQF and 6Y2E)	Catalytic pocket		CA Tyr 37	30 x 30 x 30 Å
MPro Dimer **(PDB Ids: 6WQF & 6Y2E)**	protein-protein interface These residues were close to the two cavities/pocket identified by the MOE site finder in the PPI of the dimer	Chain1: SER1 GLY2 PHE3 ARG4 LYS5 MET6 ALA7 VAL125 TYR126 GLN127 CYS128 ALA129 ARG131 THR135 ILE136 LYS137 GLY138 SER139 GLY170 VAL171 HSD172 ASP197 THR198 THR199 VAL204 TRP207 ILE213 ASN214 ASP216 TRP218 TYR239 ILE281 LEU282 GLY283 SER284 ALA285 LEU286 LEU287 GLU288 ASP289 GLU290 PHE291 Chain 2:PHE3 ARG4 LYS5 MET6 ALA7 VAL125 TYR126 GLN127 CYS128 ALA129 ARG131 THR135 ILE136 LYS137 GLY138 SER139 VAL171 ALA193 ALA194 GLY195 THR196 ASP197 THR198 THR199 VAL204 TRP207 ASN214 TYR237 ASN238 TYR239 LEU272 LEU282 GLY283 SER284 ALA285 LEU286 LEU287 GLU288 ASP289 GLU290 PHE291	CA TYR 126 (chain A)	40 x 40 x 40 Å
		Chain1: GLY11 LYS12 VAL13 GLU14 LYS97 PRO99 LYS100 ASP155 CYS156 VAL157 PHE305 GLN306 Chain2: GLY11 LYS12 GLU14 GLY15 MET17 VAL18 GLN19 GLN69 ALA70 GLY71 LYS97 ASN119 GLY120 SER121 PRO122	CA SER 10 (chain A)	40 x 40 x 40 Å

NSP9 Monomer **(PDB Id: 6W4B)**	Protein-Protein Interface	6 to 9, 97 98 100 102 104 105 106 108 109 110 112	2.99, 15.23, 2.21	112 x 36 x 76 Å
NSP9 Dimer **(PDB Id: 6W4B)**	Whole protein		CA MET 102	40 x 40 x 40 Å
NSP15 Monomer **(PDB Id: 6VWW)**	Large region including the catalytic pocket		CA VAL 166	70 x 40 x 40 Å
NSP15 Dimer **(PDB Id: 6VWW)**	Centered on hydrophobic core of the protein and the protein-protein interface		CA ASN 1413 (chain E)	70 x 50 x 50 Å
PLPro **(PDB Id: 6W9C)**	Dimer	LEU160 GLY161 ASP162 VAL163 ARG164 MET206 SER243 ALA244 PRO245 PRO246 TYR262 TYR266 GLN267 CYS268 TYR271 THR299 ASP300 SER101 CYS109 ALA112 THR113 LEU116 HSP270 TYR271 LYS272 HSD273 ILE283 ASP284 GLY285 TYR 265 TYR 269 ASP 165	CA TYR 262	40 x 40 x 40 Å

NSP3 Asymmetric Unit (monomer extracted) (PDB Id: 6W02)	Catalytic pocket	21 to 24, 44 to 52, 38 to 40, 125 to 133, 154 to 160	-18.67, 1.68, 10.08	20 x 24 x 12 Å
S Protein RBD Apo **PDB Id: 6W41**	Spike-ACE2 Interface	613 614 615 636 637 638 639 664 665 666 667 668 669 670 671 672 673 674 678 679 680 681 682 683 684 685 686 687 688 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773		
S Protein RBD with ACE2 PDB: 6W41	Spike ACE2 Interface	0 1 10 11 12 13 14 15 16 17 18 19 2 20 21 22 23 24 25 26 27 28 29 290 293 3 30 302 303 304 305 306 307 308 309 31 310 311 312 313 314 315 317 32 321 328 329 33 330 331 332 333 334 335 336 337 338 339 34 340 357 359 36 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 38 39 4 40 41 42 43 44 45 46 47 48 49 5 50 51 52 53 536 539 54 540 541 542 543 544 55 56 57 58 59 6 60 61 613 614 615 62 63 636 637 638 639 64 65 66 664 665 666 667 668 669 67 670 671 672 673 674 678 679 68 680 681 682 683 684 685 686 687 688 69 7 70 700 701 702 703 704 705 706 707 708 709 71 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 73 731 732 733 734 735 736 737 738 739 74 740 741 742 743 744 745 746 747 748 749 75 750 751 752 753 754 755 756 757 758 759 76 760 761 762 763 764 765 766		

		767 768 769 77 770 771 772 773 78 79 8 80 81 82 83 84 87 9		
NSP3 phosphatase domain (PDB: 6W02)	Phophatase-active site	GLY48, VAL49, LEU122, LEU123, ALA124, PRO125, LEU126, ILE131, PHE132		

Table S3. Identity between compounds predicted to be in the top 500 lists from clusters derived from the first 100ns of T-REMD clusters for nine selected systems.

target	**overlap in top 500 (%)**
Nsp15, endoribonuclease hexamer	62.4
Nsp15, endoribonuclease monomer	59.4
MPro dimer interface	55.4
Nsp10/Nsp16 complex patched	45.2
Nsp9 dimer	56.8
N-protein tetramer	67.2
PLPro (neutral)	51.2
S-protein apo	60
S-protein Ace-2 complex	46.8
N-protein	61.6

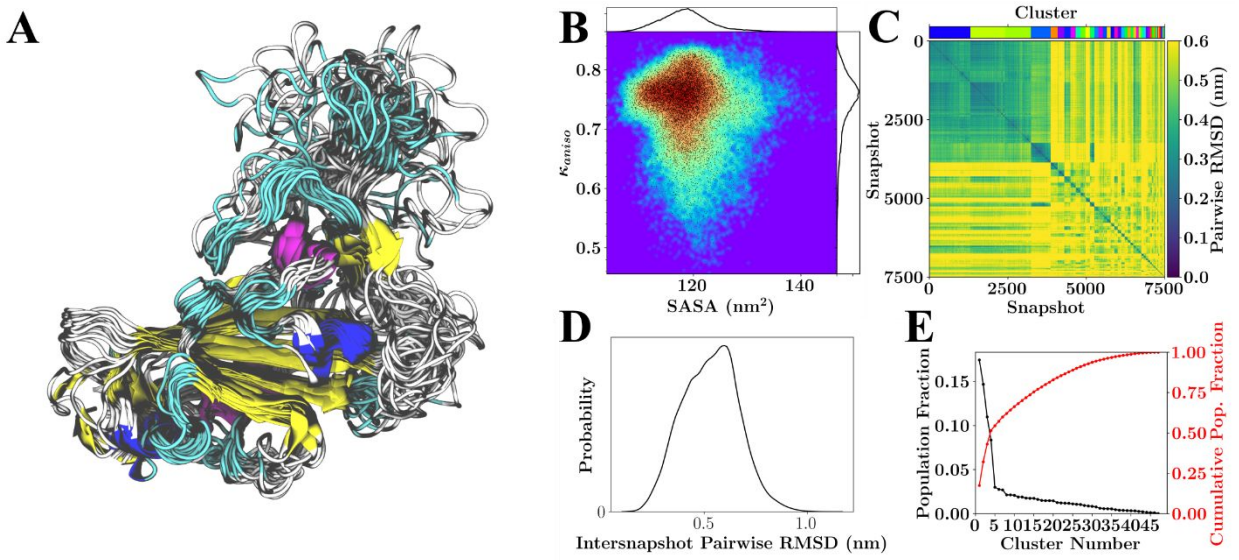

Figure S1. General results for simulations of S Protein RBD without ACE2. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

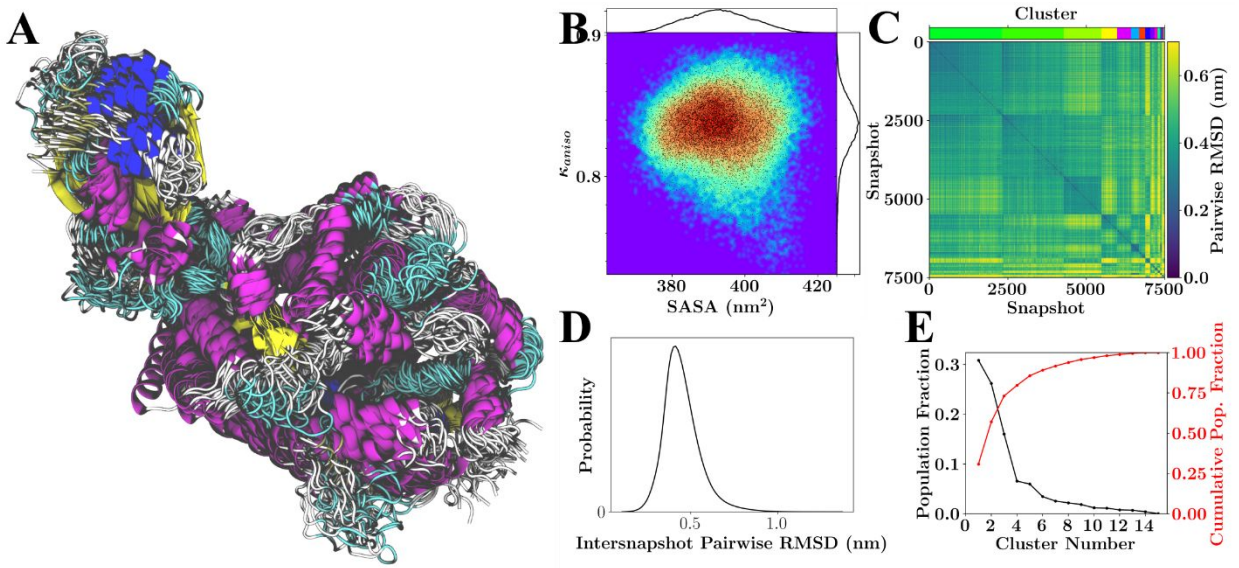

Figure S2. General results for simulations of S Protein in complex with ACE2. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

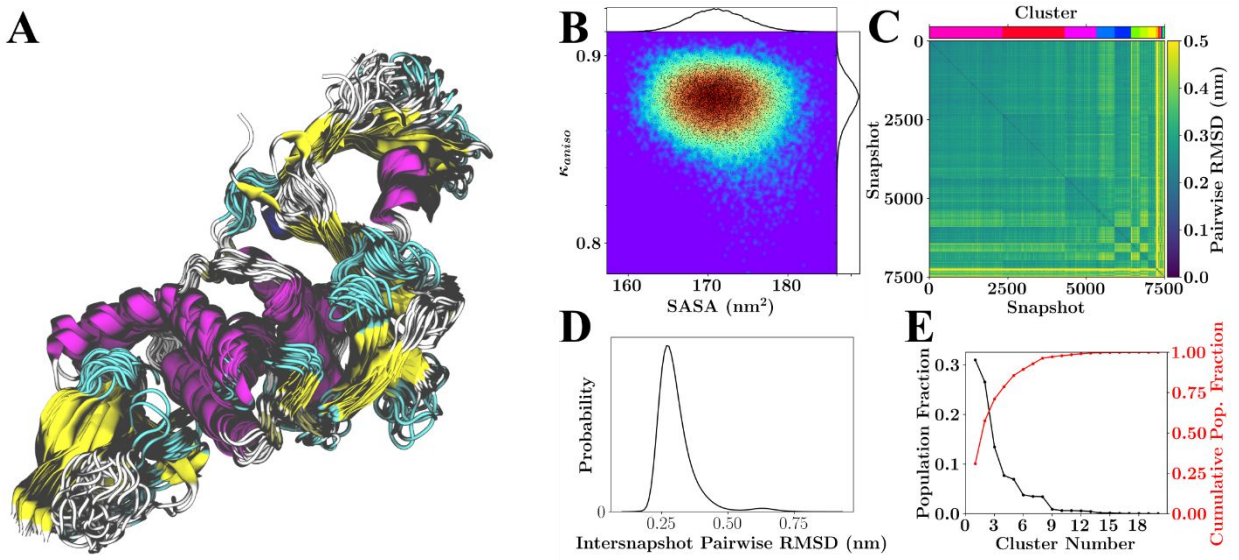

Figure S3. General results for simulations of PLPro with charged HIS protonation states. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

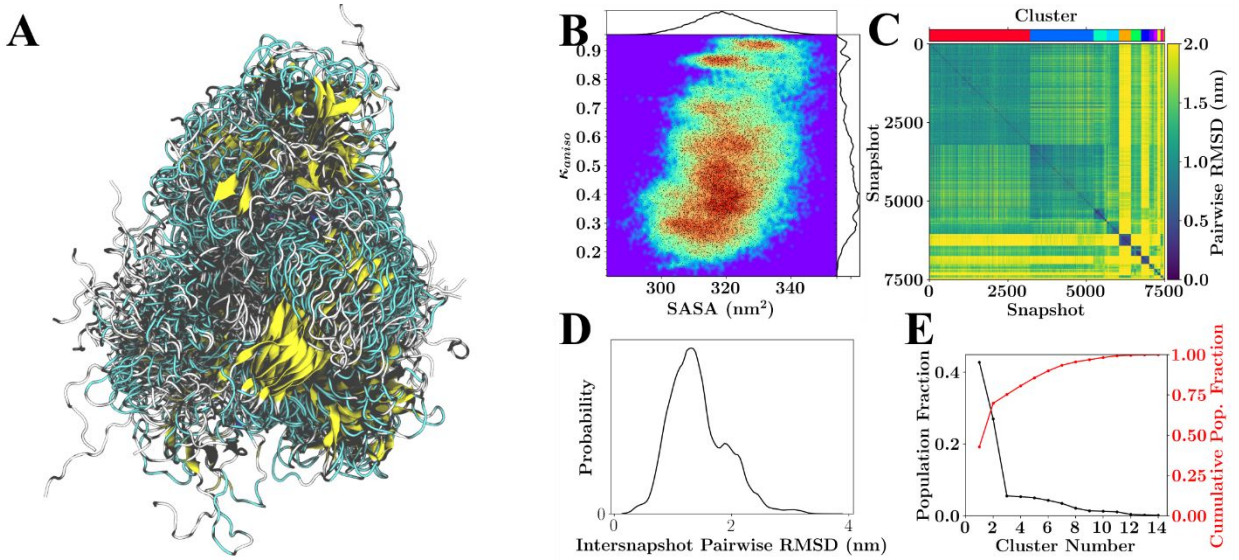

Figure S4. General results for simulations of the tetrameric complex of the N-terminal of the N Protein in the absence of Zn. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

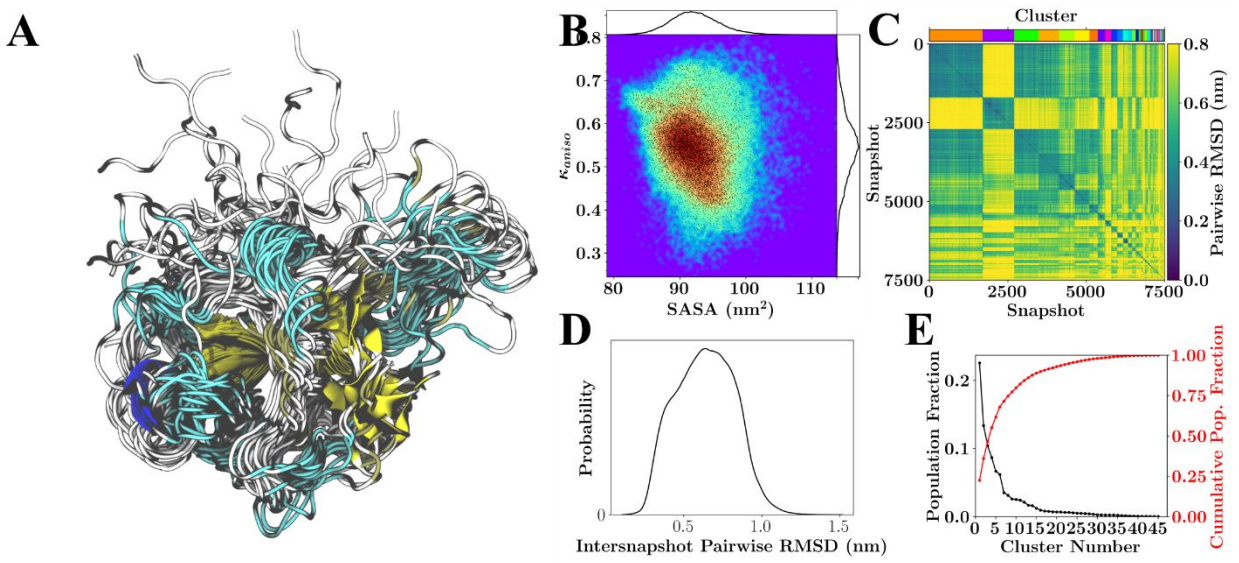

Figure S5. General results for simulations of N Protein N-terminus monomer using N Protein N-terminus without bound Zn as the initial structure. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

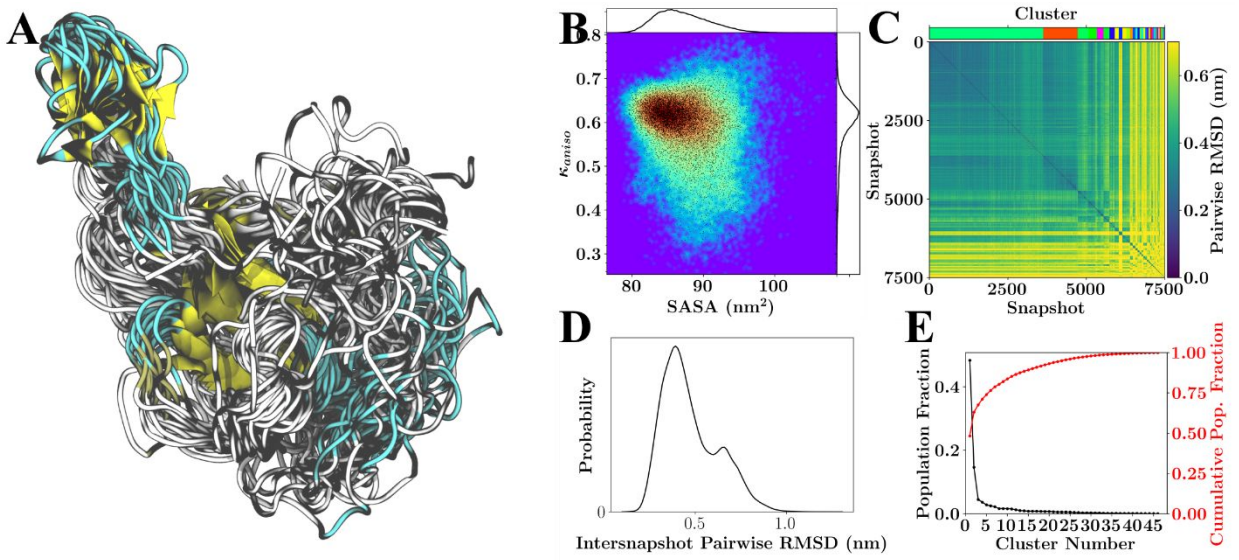

Figure S6. General results for simulations of N Protein Phosphoprotein (N-terminus) monomer states with initial conditions derived from the N Protein complex crystallized with Zn (though no Zn is bound in the monomer, derived from PDB: 6VYO). (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

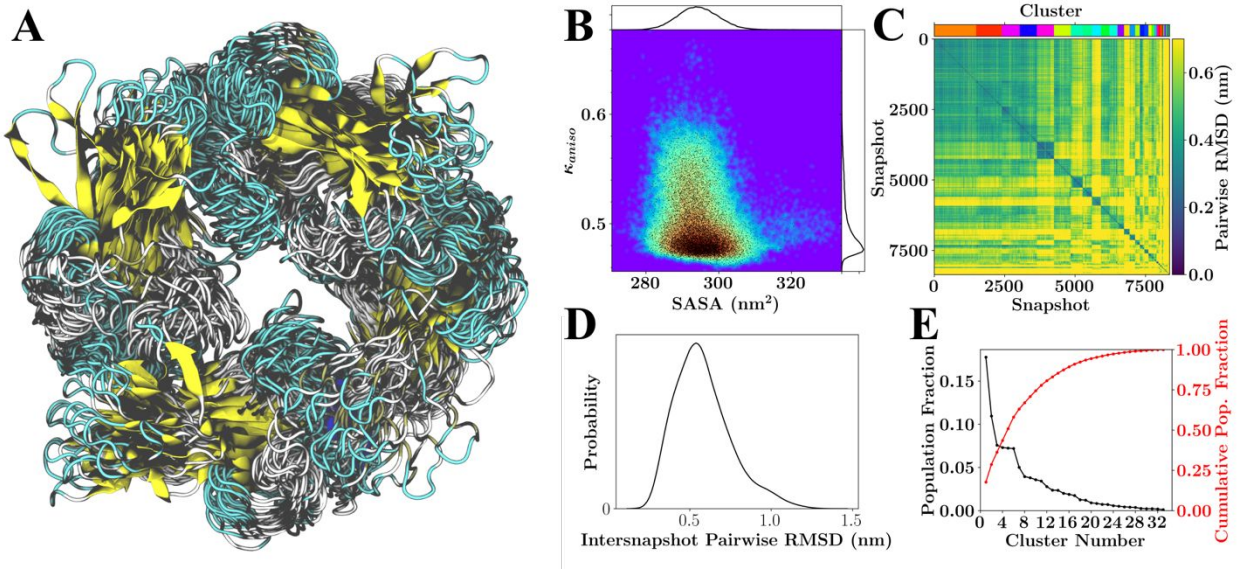

Figure S7. General results for simulations of N Protein Phosphoprotein (N-terminus) complex states with initial conditions derived from the N Protein complex crystallized in the presence of Zn (PDB 6VYO). (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

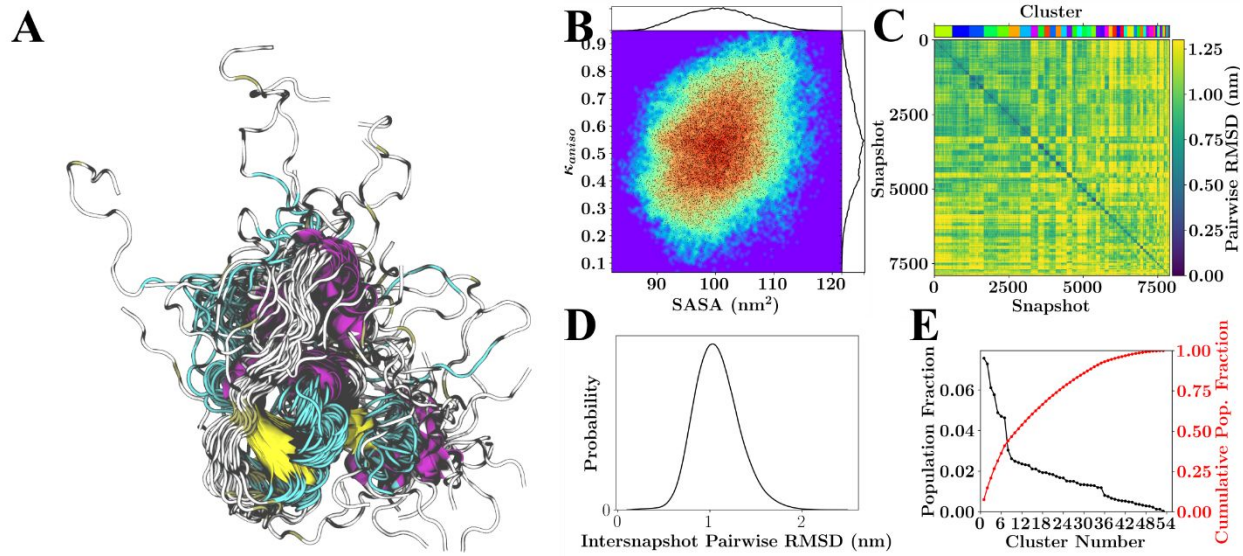

Figure S8. General results for simulations of NSP10 monomer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

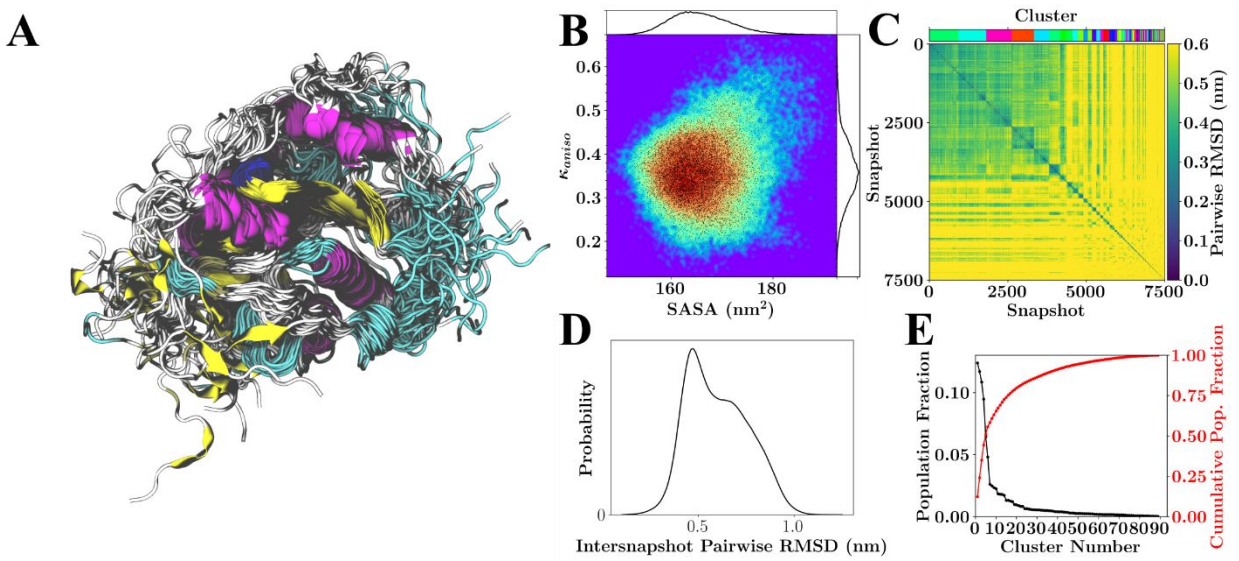

Figure S9. General results for simulations of NSP16 monomer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

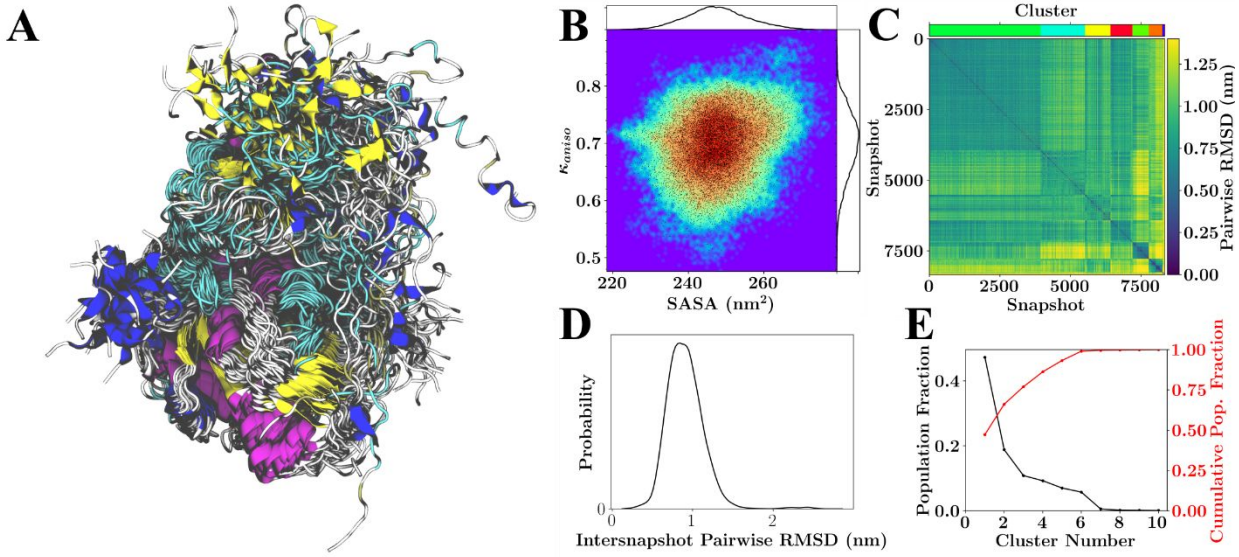

Figure S10. General results for simulations of NSP10+NSP16 Heterodimer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

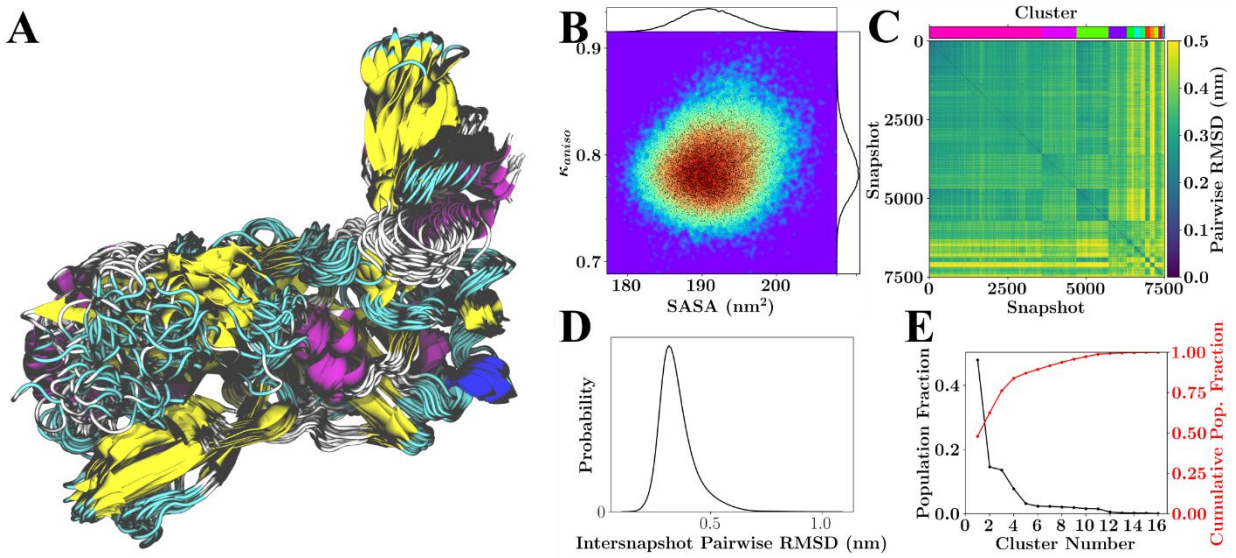

Figure S11. General results for simulations of NSP15 monomer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

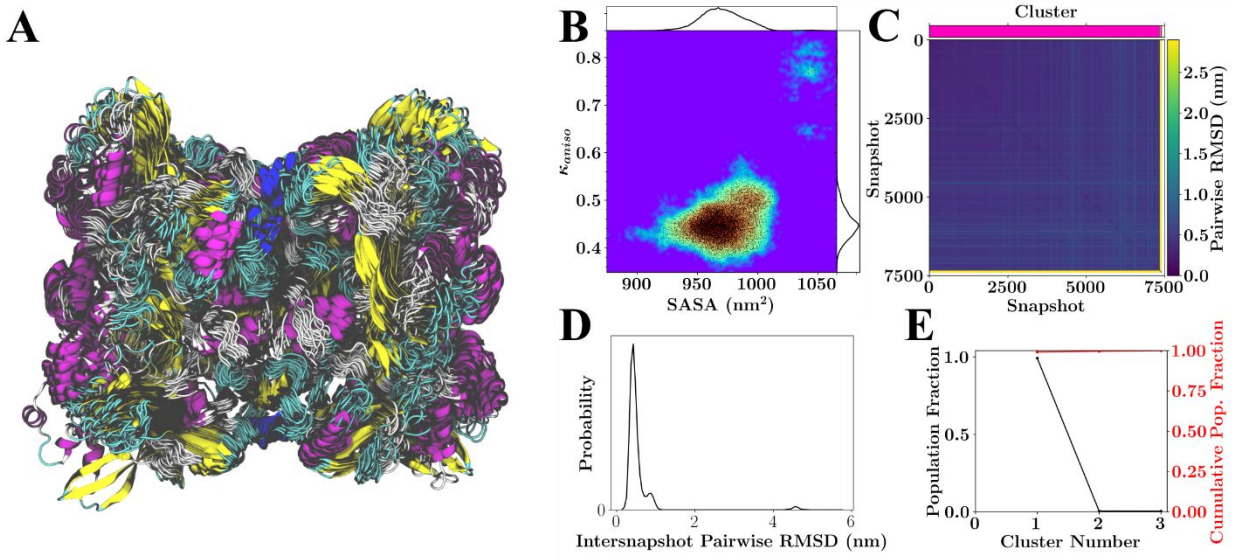

Figure S12. General results for simulations of NSP15 hexamer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

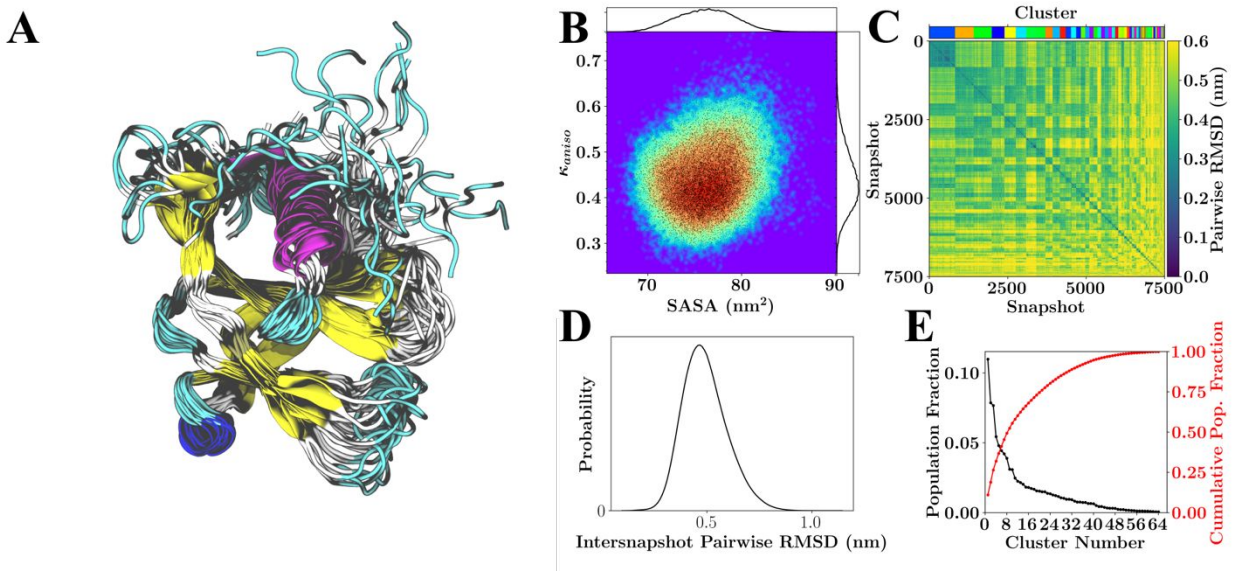

Figure S13. General results for simulations of NSP9 monomer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

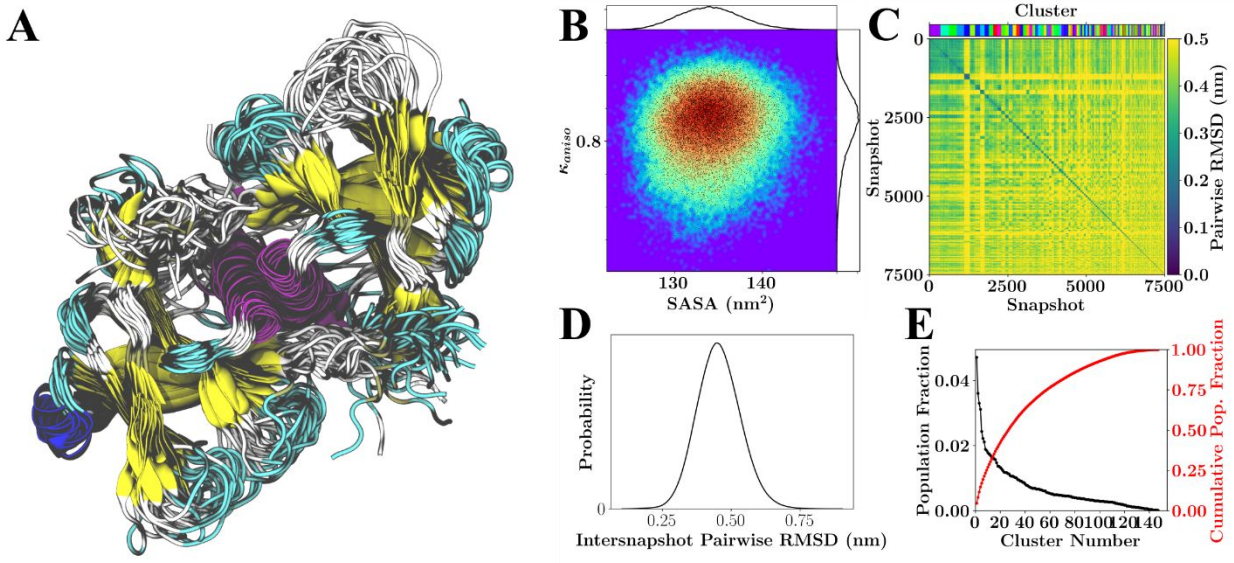

Figure S14. General results for simulations of NSP9 Dimer. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

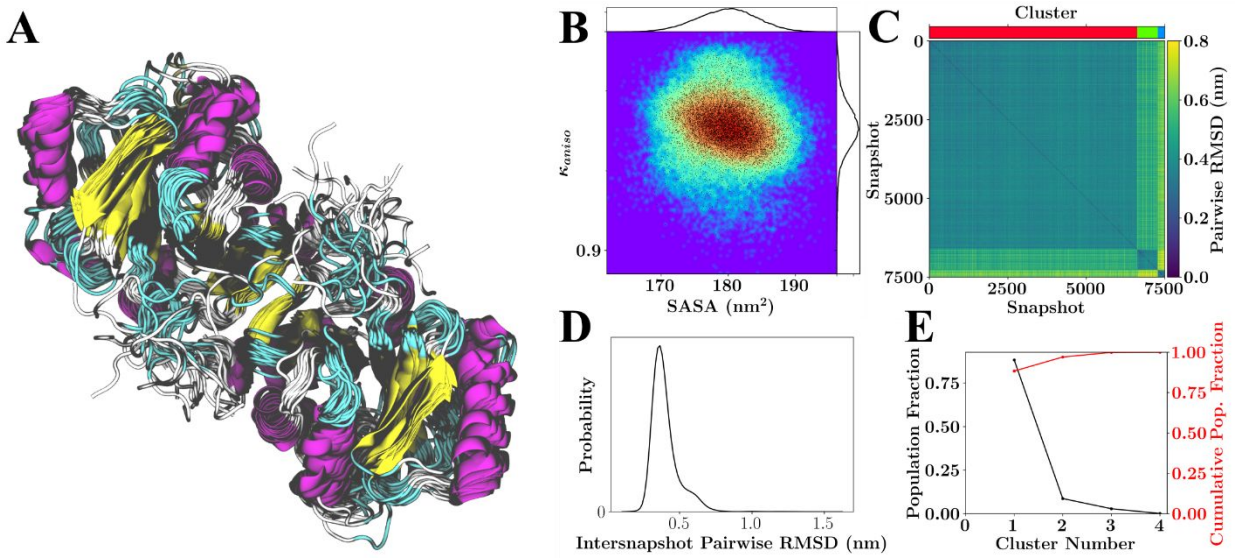

Figure S15. General results for simulations of NSP3 Phosphatase Domain asymmetric unit. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

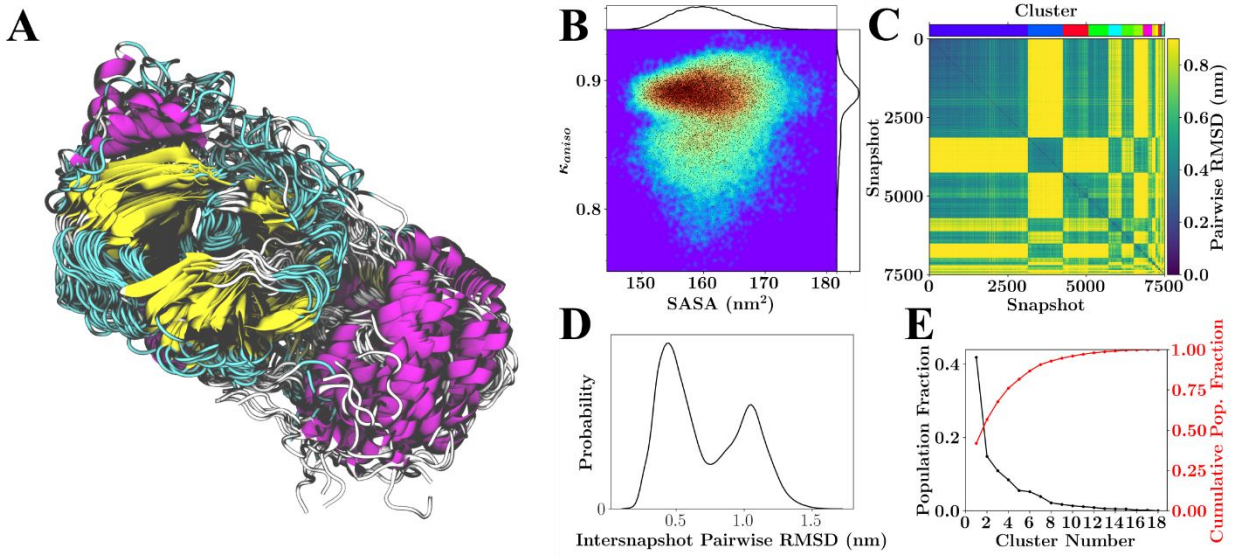

Figure S16. General results for simulations of MPro monomer with all HIS treated as HID (Default Output from CHARMM-GUI). (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

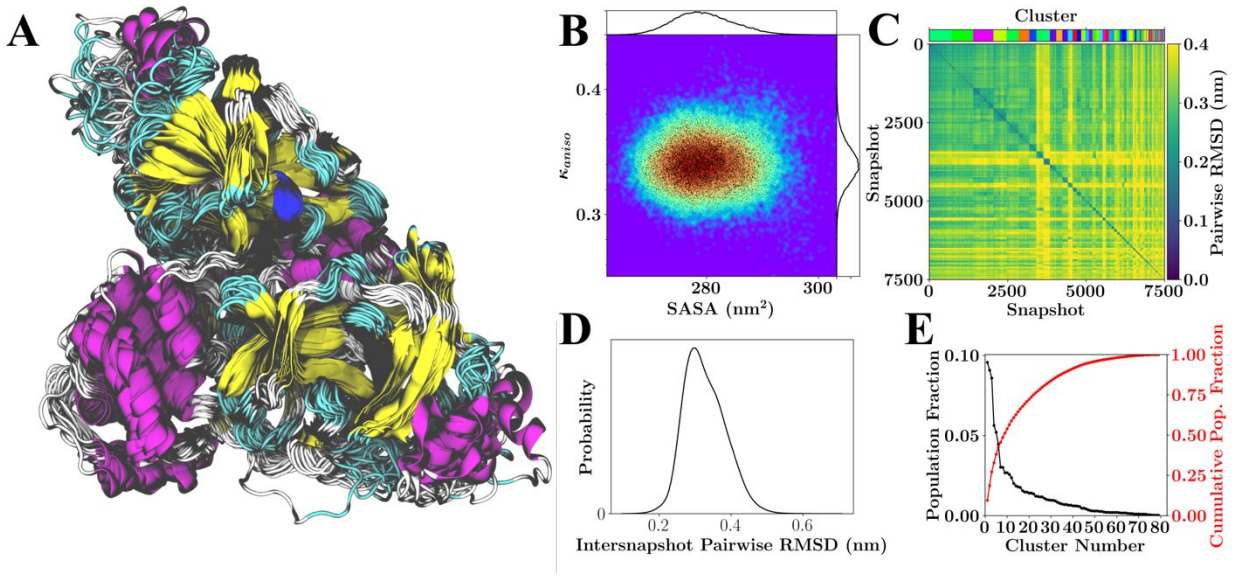

Figure S17. General results for simulations of MPro Dimer with all HIS treated as HID (Default Output from CHARMM-GUI). (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

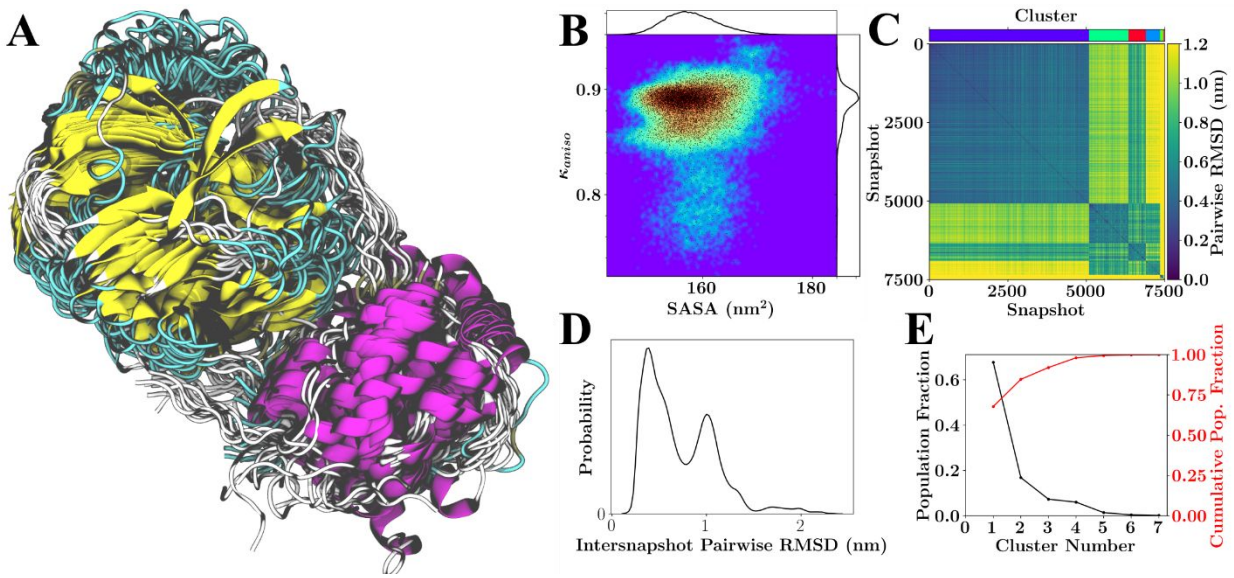

Figure S18. General results for simulations of MPro monomer with the following protonation states HSD41, HSD64, HSD80, HSE163, HSE164, HSE172, & HSE246. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

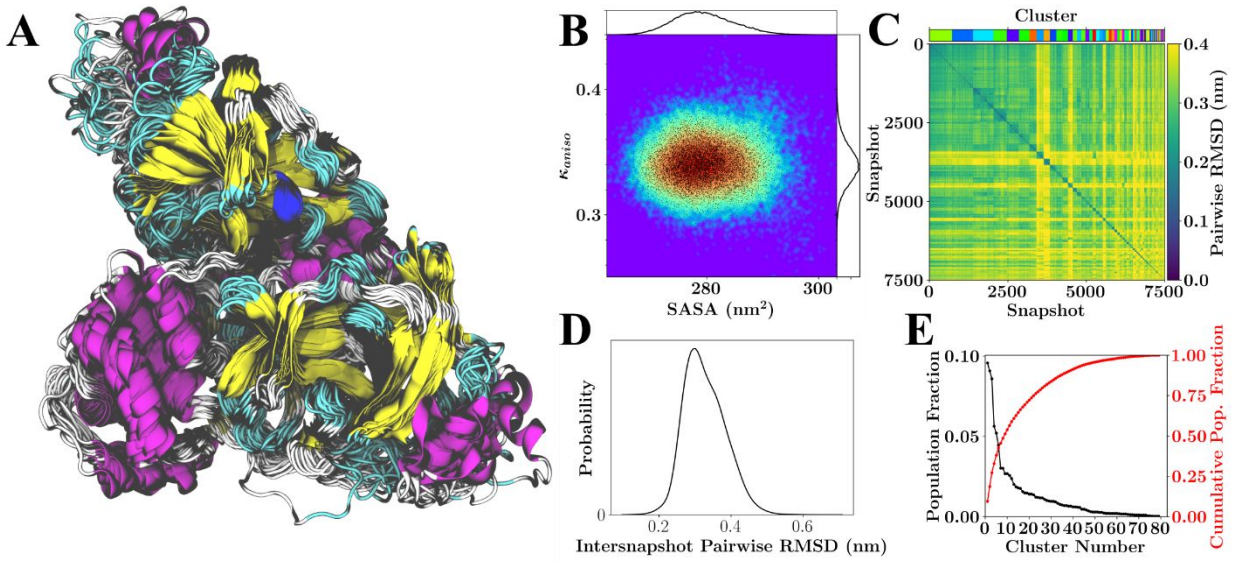

Figure S19. General results for simulations of MPro dimer with the following protonation states HSD41, HSD64, HSD80, HSE163, HSE164, HSE172, & HSE246. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

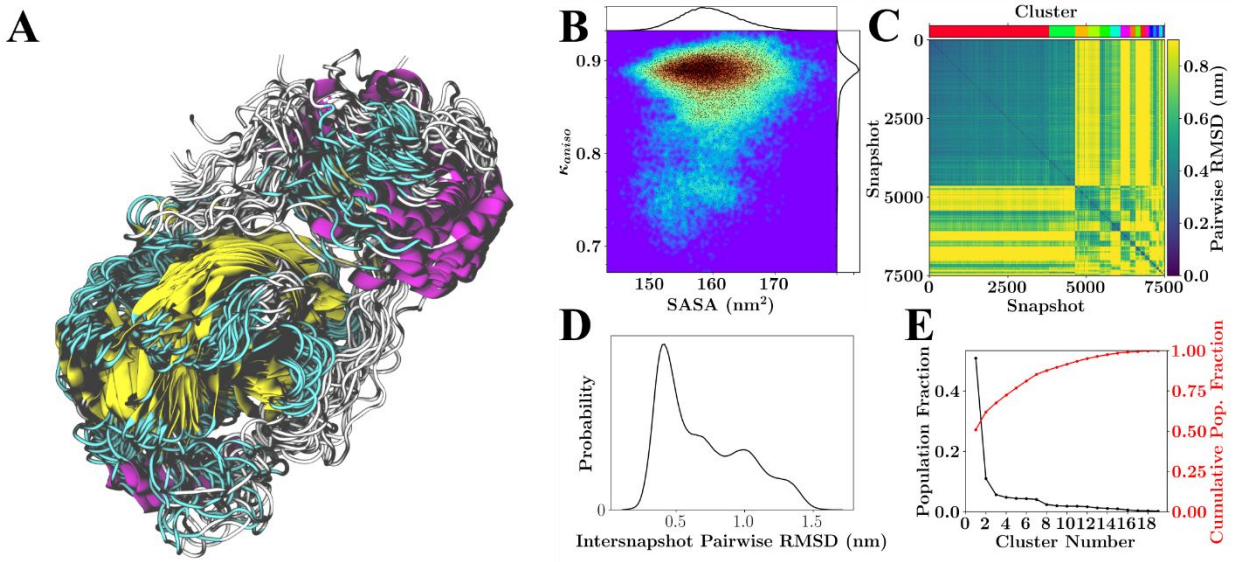

Figure S20. General results for simulations of MPro monomer with the following protonation states HSE41, HSD64, HSD80, HSE163, HSE164, HSE172, & HSE246. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

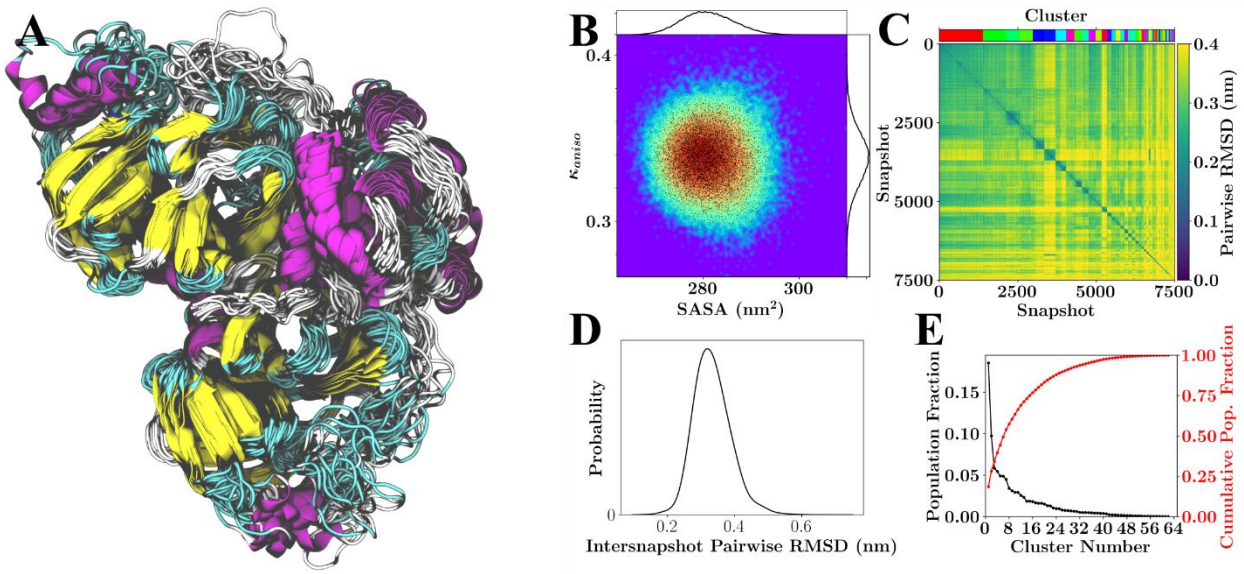

Figure S21. General results for simulations of MPro monomer with the following protonation states HSE41, HSD64, HSD80, HSE163, HSE164, HSE172, & HSE246. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

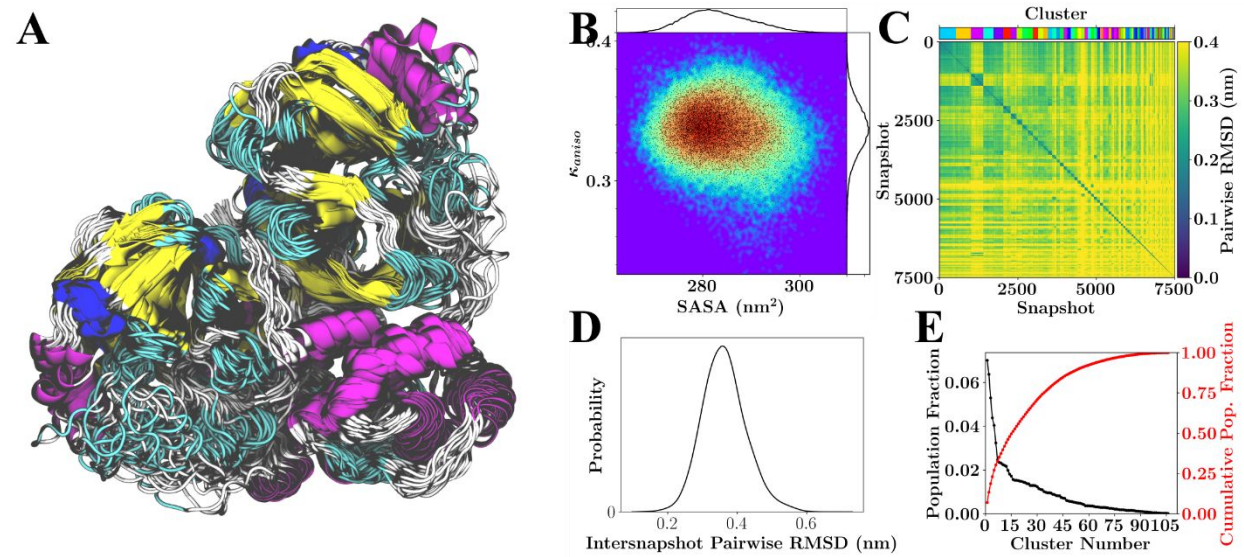

Figure S22. General results for simulations of the 'charged' MPro dimer with the following protonation states HSD41, HSP64, HSP80, HSP163, HSP164, HSE172, HSE246. (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

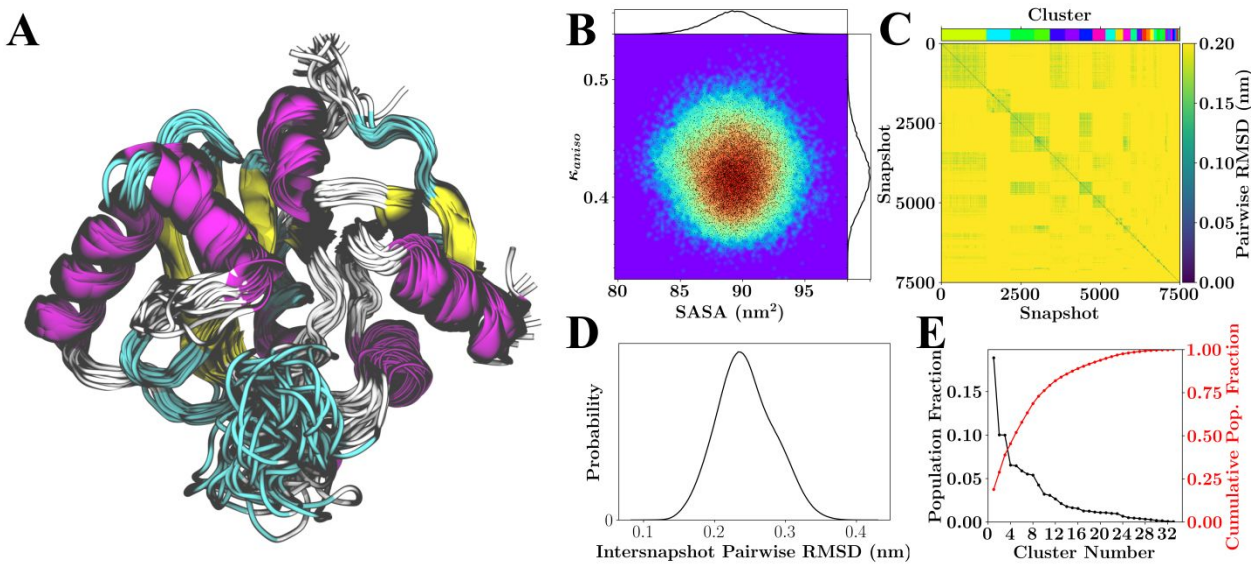

Figure S23. General results for simulations of the NSP3 X-Domain (Phosphatase) (A) Overlay of 26 RMSD aligned structures from the lowest temperature replicate spanning the 750 ns of sampling. (B) Population distribution for shape anisotropy (κ) and solvent accessible surface area (SASA), with redder colors indicating greater occupancy of these kappa-SASA combinations. The distributions are also reflected by one-dimensional histograms above and to the right of the plot, and black dots within the population distribution, which represent position information for 10% of the total snapshots considered. (C) Pairwise RMSD clustering for the lowest temperature replica, with the snapshots ordered according to their cluster. The clusters in this instance were defined using a cutoff of half the maximum RMSD observed within the simulation and are labeled according to color with a color-bar for reference located above the plot. (D) Pairwise RMSD distribution across all snapshots. (E) Population statistics for the clusters introduced in (C).

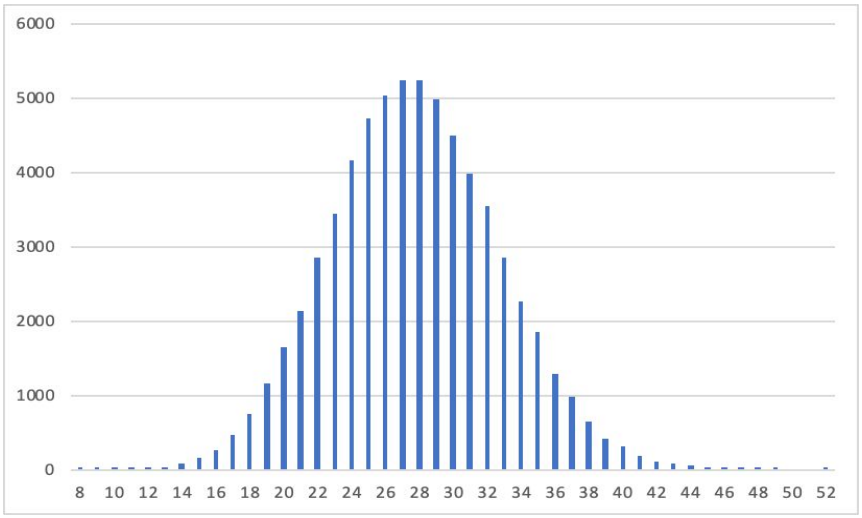

Figure S24. Distribution of expected numbers of identical compounds for two random hits from 500 compounds out of the complete smaller database.

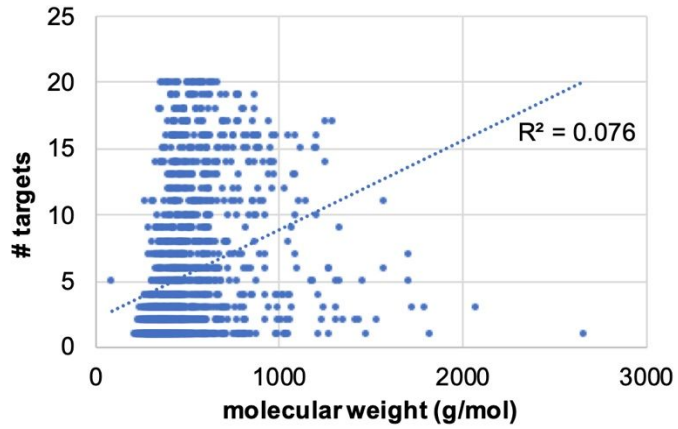

Figure S25. Correlation of ligand molecular weight and the number of targets for which a particular ligand was in the top 500 ranked compounds.