Supplementary Information

Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease

Chinmayee Choudhury[1*]

[1]Department of Experimental Medicine and Biotechnology, PGIMER, Sector 12, Chandigarh-160012, India. Email: chinmayee.choudhury@gmail.com

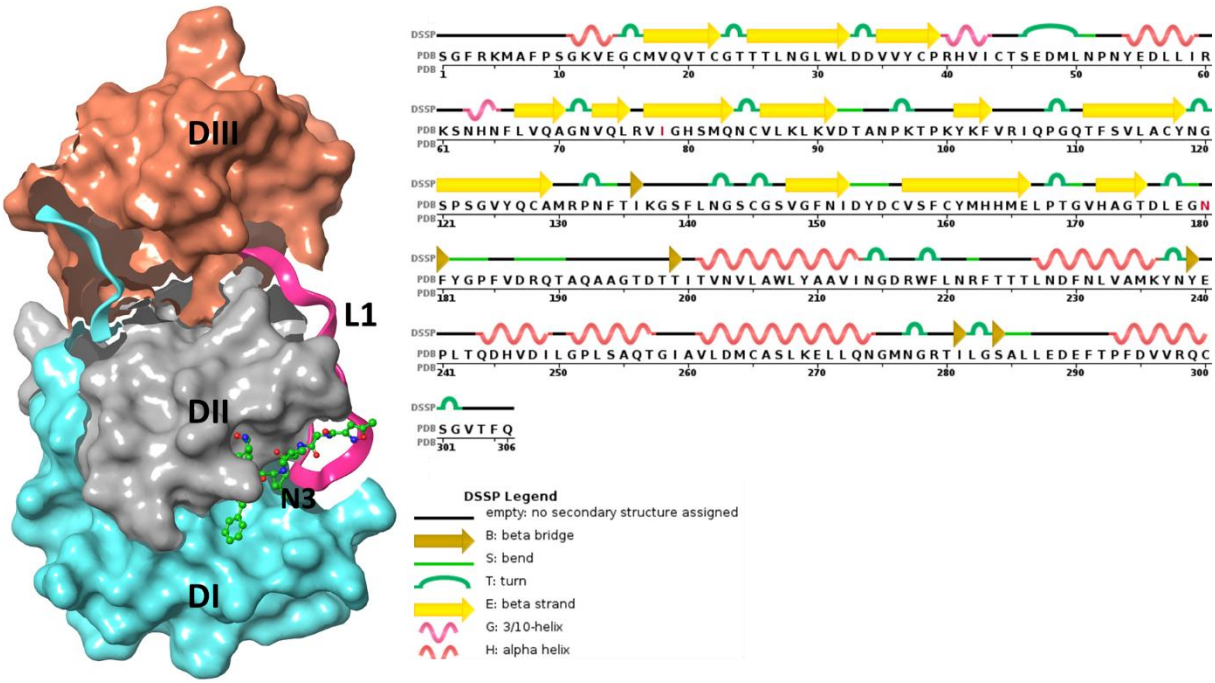

Figure S1. the domains and the overall secondary structure of the nCov-MP (6LU7)

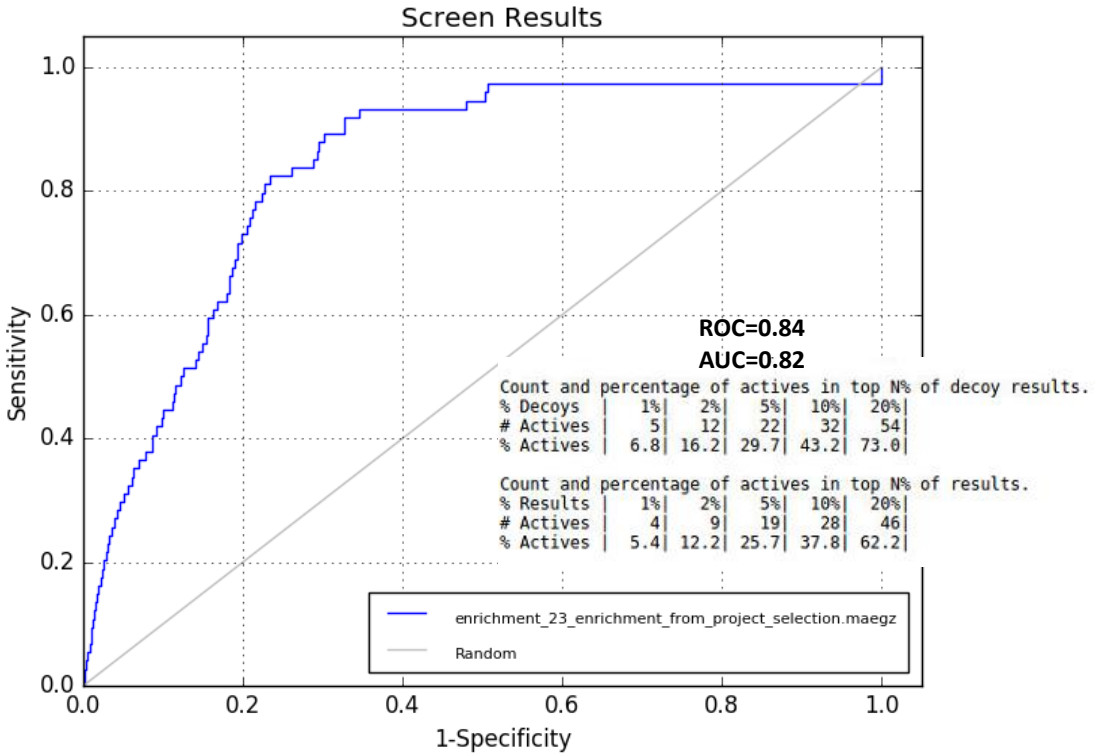

Figure S2. ROC curve showing the docking enrichment analysis.

List S1. Details of all types QuickProp descriptors calculated in this study (recommended range of value for each property is mentioned in brackets).

- **#stars** Number of property or descriptor values that fall outside the 95% range of similar values for known drugs. The following properties and descriptors are included in the determination of #stars: MW, dipole, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, volume, #rotor, donorHB, accptHB, glob, QPpolrz, QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, logS, QPLogKhsa, QPlogBB, #metabol (0 – 5)
- **#rotor** Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds. (0 – 15)
- **mol_MW** Molecular weight of the molecule. (130.0 – 725.0)
- **SASA** Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius. (300.0 – 1000.0)
- **FOSA** Hydrophobic component of the SASA (saturated carbon and attached hydrogen). (0.0 – 750.0)
- **FISA** Hydrophilic component of the SASA (SASA on N, O, and H on heteroatoms). (7.0 – 330.0)
- **PISA (**carbon and attached hydrogen) component of the SASA. (0.0 – 450.0)
- **WPSA** Weakly polar component of the SASA (halogens, P, and S). (0.0 – 175.0)
- **volume** Total solvent-accessible volume in cubic angstroms using a probe with a 1.4 Å radius. (500.0 – 2000.0)
- **donorHB** Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer. (0.0 – 6.0)
- **accptHB** Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, sothey can be non-integer. (2.0 – 20.0)
- **glob** Globularity descriptor. (0.75 – 0.95)
- **QPpolrz** Predicted polarizability in cubic angstroms. (13.0 – 70.0)
- **QPlogPC16** Predicted hexadecane/gas partition coefficient. (4.0 – 18.0)
- **QPlogPoct‡** Predicted octanol/gas partition coefficient. (8.0 – 35.0)
- **QPlogPw** Predicted water/gas partition coefficient. (4.0 – 45.0)
- **QPlogPo/w** Predicted octanol/water partition coefficient. (–2.0 – 6.5)
- **QPlogS** Predicted aqueous solubility, log S. S in mol dm–3 is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid. (–6.5 – 0.5)
- **CIQPlogS** Conformation-independent predicted aqueous solubility, log S. S in mol dm–3 is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid. –6.5 – 0.5
- **QPlogHERG** Predicted IC50 value for blockage of HERG K+ channels. (concern below –5)
- **QPPCaco** Predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells are a model for the gut-blood barrier. QikProp predictions are for non-active transport. (<25 poor, >500 great)
- **QPlogBB** Predicted brain/blood partition coefficient. (–3.0 – 1.2)
- **QPPMDCK** Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the bloodbrain barrier. QikProp predictions are for non-active transport. (<25 poor, >500 great)
- **QPlogKp** Predicted skin permeability, log *K*p. (–8.0 – –1.0)
- **QPlogKhsa** Prediction of binding to human serum albumin. (–1.5 – 1.5)
- **HumanOralAbsorption** Predicted qualitative human oral absorption: 1, 2, or 3 for low, medium, or high.
- **PercentHuman- OralAbsorption** Predicted human oral absorption on 0 to 100% scale. The prediction is based on a quantitative multiple linear regression model. This property usually correlates well with HumanOralAbsorption, as both measure the same property. (>80% is high ,<25% is poor)
- **PSA** Van der Waals surface area of polar nitrogen and oxygen atoms. (7.0 – 200.0)
- **RuleOfFive** Number of violations of Lipinski's rule of five. The rules are: mol_MW < 500, QPlogPo/w < 5, donorHB ≤ accptHB ≤10. Compounds that satisfy these rules are considered druglike. (maximum is 4)

Table S1. Various components of the XP docking score and MM-GBSA binding energy.

Name	XP GScore	XP HBond	XP Electro	XP Lipophilic EvdW	MMGBSA dG Bind	MMGBSA dG Bind Coulomb	MMGBSA dG Bind Hbond	MMGBSA dG Bind vdW	MMGBSA dG Bind Lipo	Prime MMGBSA ligand efficiency
MP-In1	-10.68	-1.53	-1.19	-6.51	-80.97	-48.08	-1.72	-61.70	-21.00	-2.13
MP-In2	-10.61	-2.40	-2.45	-5.95	-80.02	-52.00	-4.89	-58.19	-17.72	-1.95
MP-In3	-11.03	-2.22	-2.63	-5.55	-79.11	-50.14	-2.86	-58.08	-19.92	-2.20
MP-In4	-11.10	-2.36	-1.30	-5.56	-76.54	-58.53	-4.20	-54.76	-16.79	-2.39
MP-In5	-11.14	-1.90	-1.03	-6.93	-76.29	-45.05	-3.08	-61.31	-21.45	-2.12
MP-In6	-11.57	-2.55	-1.70	-4.71	-75.94	-79.68	-4.40	-46.23	-21.33	-2.45
MP-In7	-11.90	-3.42	-1.39	-6.07	-75.92	3.82	-3.53	-63.89	-25.57	-1.81
MP-In8	-10.35	-1.67	-2.00	-6.32	-75.61	-1.37	-3.25	-77.24	-25.26	-2.04
MP-In9	-10.73	-1.77	-1.95	-6.45	-75.57	-48.17	-5.07	-53.74	-22.53	-1.99
MP-In10	-11.29	-2.52	-1.85	-5.98	-74.69	-69.20	-2.92	-47.31	-20.99	-2.41
MP-In11	-11.46	-2.99	-1.61	-5.36	-74.01	-61.58	-3.91	-49.53	-19.09	-2.39
MP-In12	-10.10	-1.56	-1.18	-6.17	-72.28	-1.17	-1.58	-64.71	-25.42	-1.95
MP-In13	-11.11	-1.88	-2.68	-4.78	-71.68	-59.21	-2.40	-62.89	-14.86	-1.75
MP-In14	-11.63	-1.89	-1.39	-5.60	-71.12	-90.87	-2.97	-63.15	-15.81	-1.65
MP-In15	-10.05	-1.67	-1.50	-5.71	-70.85	-4.01	-1.87	-67.26	-26.38	-1.73
MP-In16	-11.68	-2.92	-1.49	-5.73	-70.42	-57.54	-3.53	-47.14	-19.55	-2.52
MP-In17	-10.74	-0.86	-0.64	-7.26	-69.92	-5.22	-1.11	-62.04	-22.56	-1.84

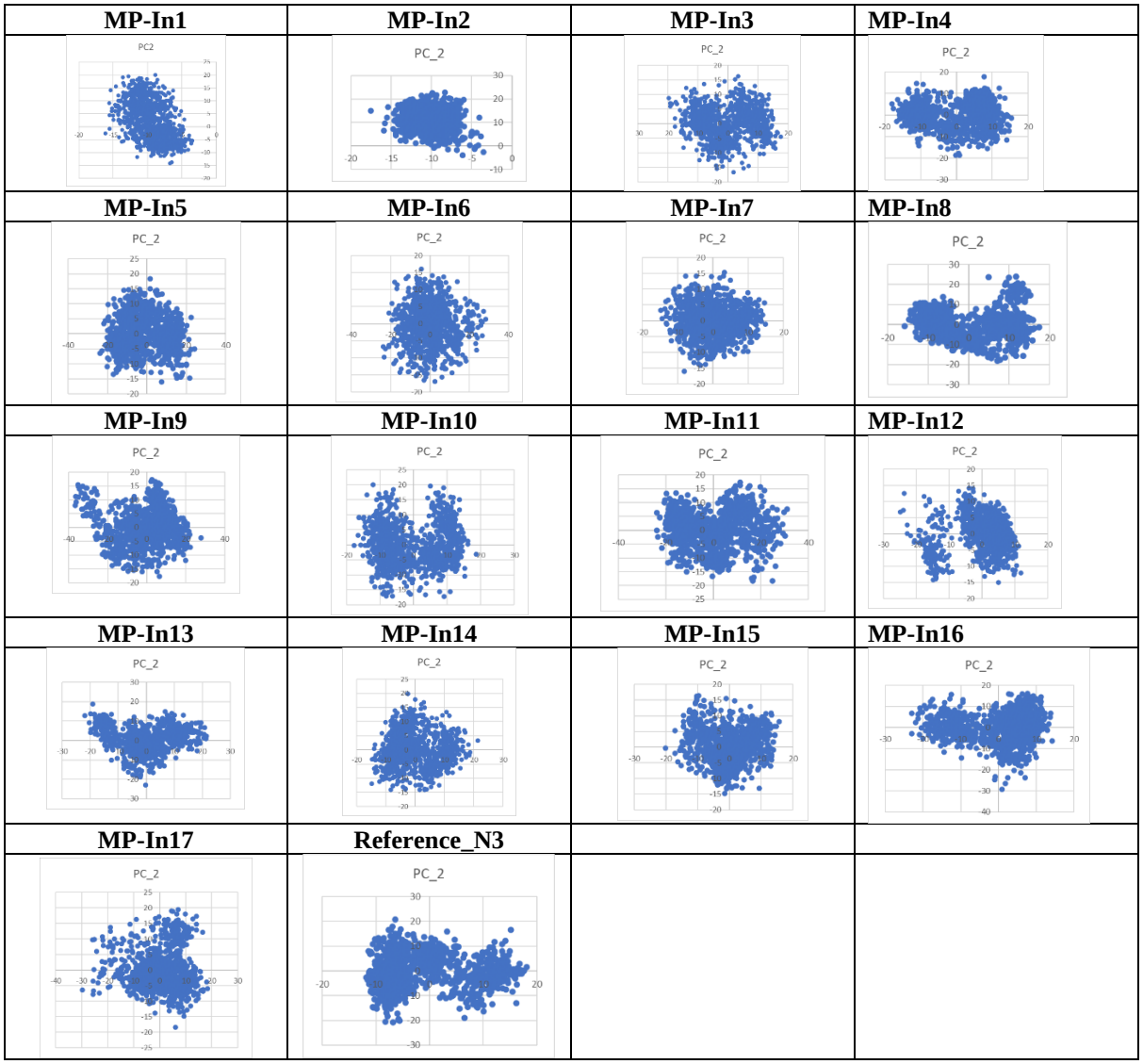

Figure S3. Projection of the motion for Cα aoms of the protein bound to the 18 ligands (including one reference structure 6LU7) in phase space along the PC1 and PC2.

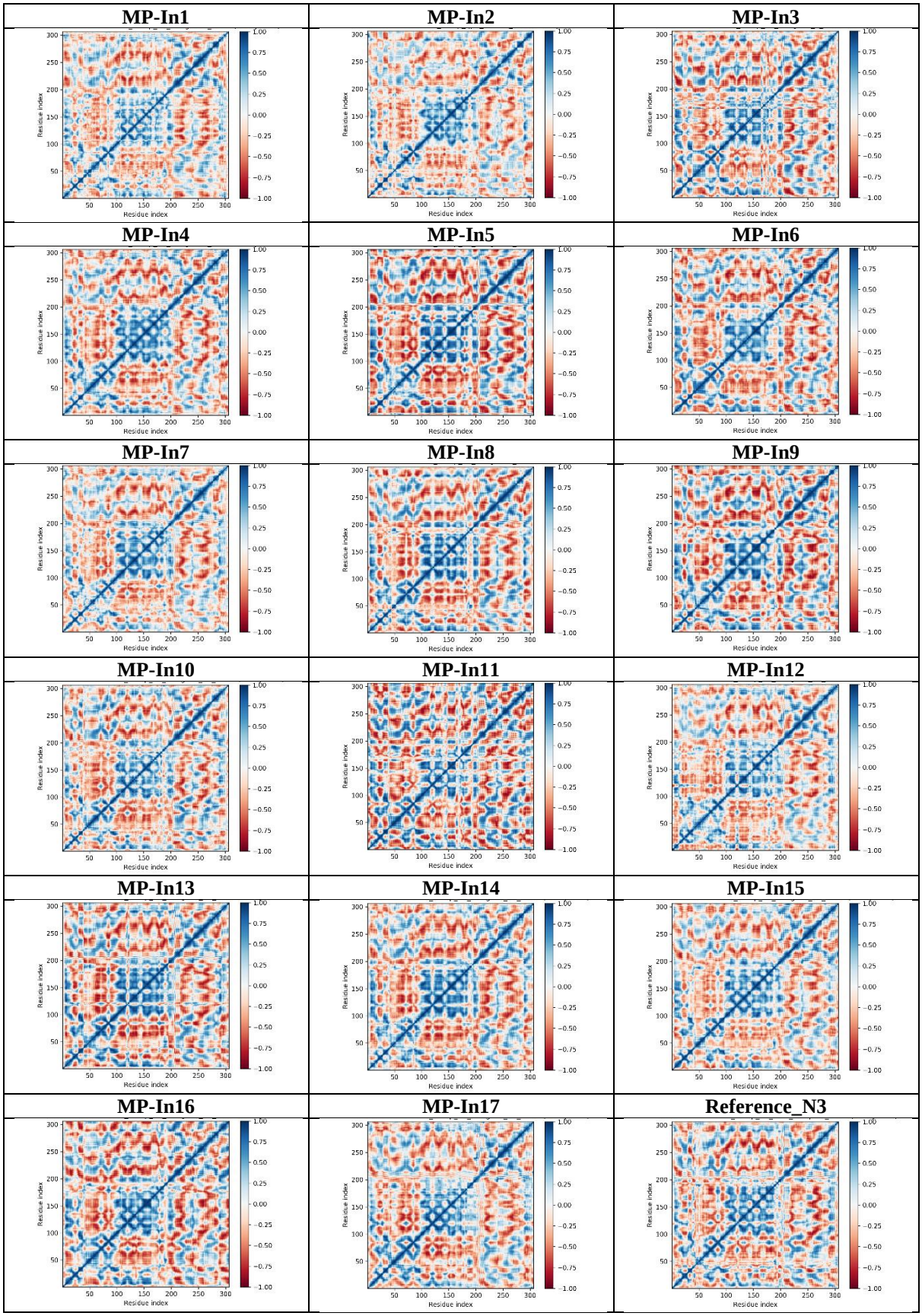

Figure S4. Cross-correlation plots for the 18 complexes

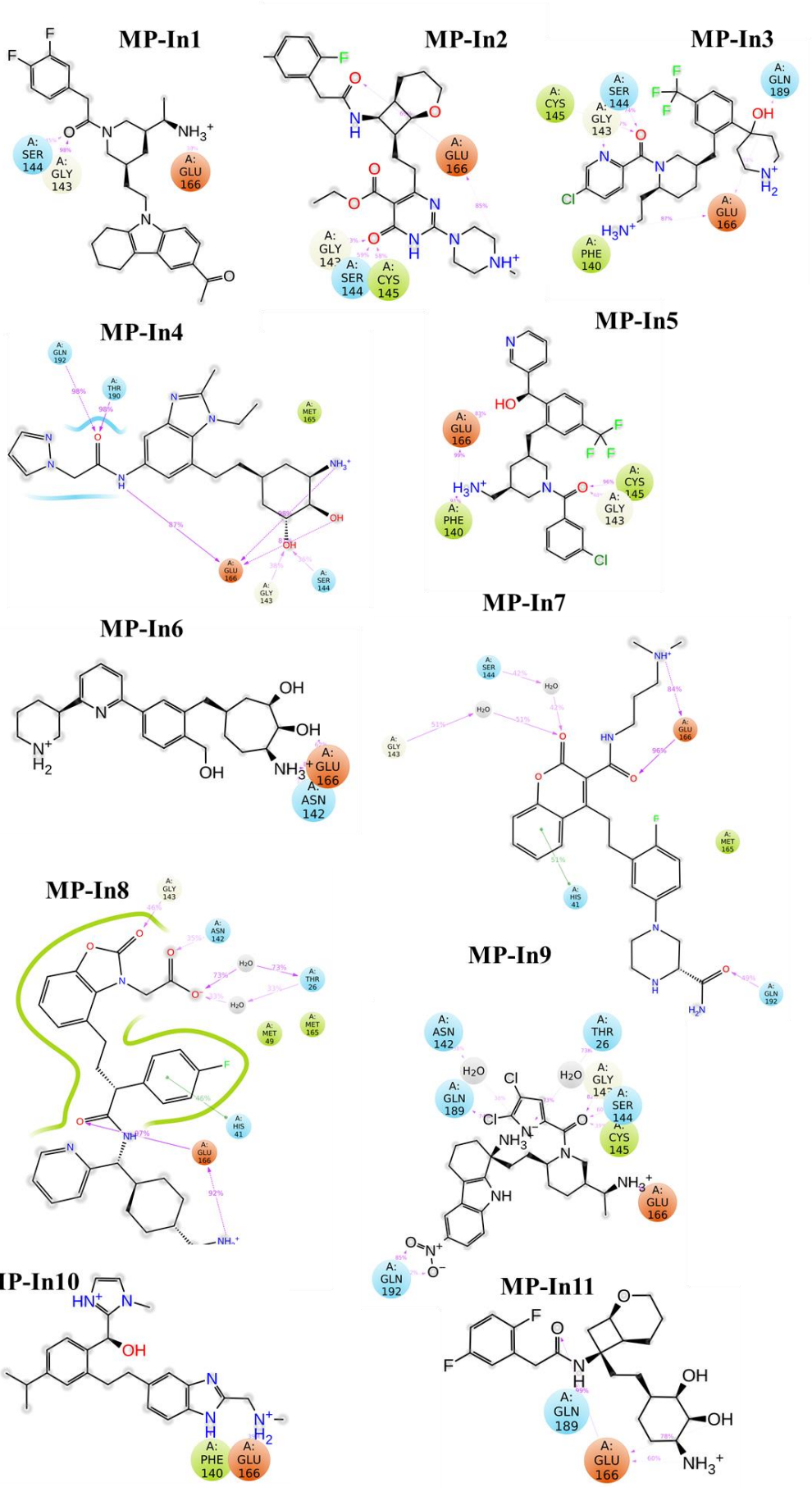

Figure continues …

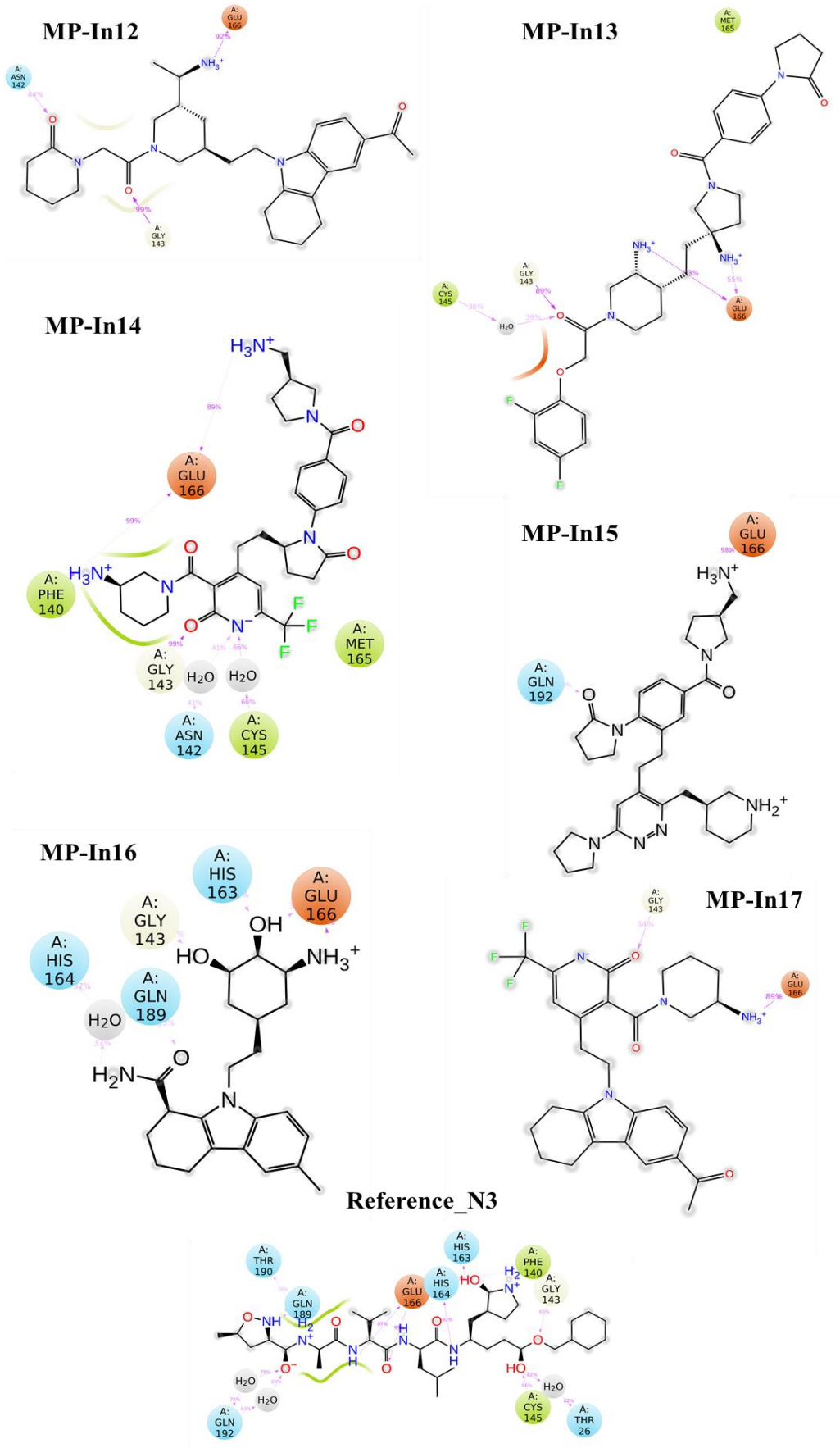

Figure S5. Interactions that occur more than 30% of the simulation time in each trajectory through 0 to 50 ns