Electronic Supplementary Information for "Characterization of the non-covalent interaction between the PF-07321332 inhibitor and the SARS-CoV-2 main protease"

Marina Macchiagodena, Marco Pagliai, and Piero Procacci*

Dipartimento di Chimica "Ugo Schiff", Universit degli Studi di Firenze, Via della Lastruccia 3, Sesto Fiorentino, I-50019 Italy

E-mail: procacci@unifi.it

Docking

In Table S1 we report the pdbcode and the binding modes in 28 co-crystal structures of $3CL^{pro}$ with noncovalently bound ligand. The crystal structures were downloaded from the PDB `ftp` site https://ftp.rcsb.org/pub/pdb/data/structures/all/pdb/. Raw PDB `.ent` files where initially processed with the unix command purging all comments and non-ligand and non-protein adducts coordinates including water.

```
for i in *ent; do\
   grep -e "^ATOM  " -e "^HETATM" $i \
   | grep -e " A[ 0-9]" | grep -v -e "[ AB]HOH[ 0-9]" \
   | grep -v " GOL " | grep -v " DMS" | grep -v " MES" \
   | grep -v "CL    CL" > $i.clean; \
```

Table S1: Binding sites of ligands noncovalently bound to 28 crystal structures of $3CL^{pro}$ (with a non zero S1 entry). See Ref.[?] for binding criteria.

PDBcode	S1′	S1	S2	S4	Interface	Surface
5r83	0.0	1.0	0.5	0.0	0.0	0.0
5r84	0.0	1.0	1.0	0.0	0.0	0.0
5re4	0.0	1.0	0.0	0.0	0.0	0.0
5reh	0.0	1.0	0.0	0.0	0.0	0.0
5rf3	0.0	1.0	0.0	0.0	0.0	0.0
5rf7	0.0	1.0	1.0	0.0	0.0	0.0
5rg1	0.0	1.0	1.0	0.0	0.0	0.0
5rgi	1.0	1.0	0.0	0.0	0.0	0.0
5rgk	0.0	1.0	0.0	0.0	0.0	0.0
5rgu	0.0	1.0	0.5	1.0	0.0	0.0
5rgv	0.0	1.0	0.0	0.0	0.0	0.0
5rgw	0.0	1.0	0.5	0.0	0.0	0.0
5rgx	0.0	1.0	0.5	0.0	0.0	0.0
5rgy	0.0	0.5	0.0	1.0	0.0	0.0
5rgz	0.0	1.0	0.5	0.0	0.0	0.0
5rh0	0.0	1.0	0.5	0.5	0.0	0.0
5rh1	0.0	1.0	0.0	0.0	0.0	0.0
5rh2	0.0	1.0	0.5	0.0	0.0	0.0
5rh3	0.0	1.0	0.5	0.0	0.0	0.0
5rh8	0.0	1.0	0.5	0.0	0.0	0.0
6m2n	0.5	0.5	1.0	0.0	0.0	0.0
6w63	1.0	1.0	1.0	0.5	0.0	0.0
6w79	1.0	1.0	1.0	0.5	0.0	0.0
7aha	0.5	0.5	0.0	0.0	0.0	0.0
7ans	0.5	0.5	0.5	0.0	0.0	0.0
7awu	0.0	1.0	0.0	0.0	0.0	0.0
7ju7	0.0	1.0	1.0	0.0	0.0	0.0
7kx5	1.0	1.0	1.0	0.5	0.0	0.0

```
done
```

In the file 7awu the `HETATM` specification for the protein residue CSO145 (S-hydroxycysteine) was set to `ATOM`.

Ligand and receptor files where generated from the `.clean` files (see above), based on the `ATOM/HETATM` specification, selecting the A-labelled coordinates in case of alternative structures using the unix commands:

```
grep HETATM $1.clean  |\
 awk '{a=substr($0,17,1); if(a=="A" || a== " ") print }'\
  > ${pdbcode}_LIG.pdb


grep -v HETATM $1.clean |\
 awk '{a=substr($0,17,1); if(a=="A" || a== " ") print }' \
> ${pdbcode}_REC.pdb
```

The `REC` and `LIG` files where fed to program OpenBabel[?] to add hydrogens appropriate for pH=7 and then to the AutoDock4 python tools `prepare_ligand4.py` and `prepare_receptor4.py` to generate the Autodock `pdbqt` files for Vina. Vina was launched using the center of mass of the ligand as a cubic box center with sidelength of 25 Å. The receptor was held fixed during docking. Results are reported in Table S2 and Table S3.

Table S2: Minimum RMSD (Å) between Vina-docked poses and experimental structures. Rank indicates the ranking of the pose yielding the minimum RMSD and $\Delta\Delta G$ (in kcal/mol) is its free energy difference with respect to the best scoring pose with free energy ΔG_{best} (in kcal/mol). Natoms is the number of heavy atoms in the ligand.

PDBcode	RMSD	Rank	$\Delta\Delta G$	ΔG_{best}	Natoms
5rf7	0.488	1	0.0	-6.9	19
5rh2	0.669	1	0.0	-6.3	18
6m2n	0.704	2	0.1	-7.2	20
6w79	1.042	1	0.0	-8.3	34
5rf3	1.109	3	0.4	-3.7	7
6w63	1.162	1	0.0	-8.2	34
5rgu	1.345	2	0.0	-6.9	22
5r84	1.502	2	0.2	-6.1	16
5rgz	1.721	1	0.0	-6.6	18
7ju7	1.803	1	0.0	-7.8	36
5rgw	1.896	3	0.3	-6.2	18
5rgi	2.159	7	0.7	-6.1	15
5rh3	2.309	4	0.2	-6.2	19
5rh8	2.469	5	0.2	-5.6	19
5rh1	2.480	1	0.0	-5.5	16
7aha	2.540	9	0.7	-4.0	8
7ans	3.080	5	0.1	-6.3	20
5rgv	3.210	6	0.4	-6.5	20
5rg1	3.227	5	0.3	-6.1	20
5rgx	3.313	1	0.0	-6.3	19
5r83	4.364	1	0.0	-5.9	16
5re4	4.488	1	0.0	-4.3	11
5rh0	5.065	1	0.0	-5.7	16
5rgy	5.085	3	0.4	-6.8	22
7kx5	5.837	1	0.0	-8.8	38
7awu	6.036	8	0.7	-5.8	20
5rgk	7.524	3	0.2	-6.4	18
5reh	7.611	6	0.9	-6.3	18

Table S3: RMSD (Å) between Vina-docked best poses and experimental structures.

PDBcode	RMSD	ΔG_{best}	Natoms
6w63	0.8449	-8.3	34
5rf7	0.9246	-6.9	19
6w79	0.9908	-8.4	34
5rgu	1.315	-6.9	22
7kx5	1.449	-8.9	38
5rh2	1.584	-6.4	18
7ju7	1.814	-8.2	36
5rh1	2.484	-5.5	16
5rgx	2.589	-6.2	19
5rf3	3.056	-3.8	7
5rgw	3.588	-6.4	18
5rgi	4.006	-6.1	15
7aha	4.074	-3.8	8
5rg1	4.110	-6.3	20
5r83	4.361	-5.9	16
5re4	4.513	-4.3	11
5rh0	5.047	-5.7	16
7ans	5.399	-6.4	20
5rgz	5.819	-6.0	18
5rh3	6.098	-6.2	19
5rh8	6.149	-6.0	19
6m2n	6.240	-7.0	20
5rgv	6.863	-6.6	20
5r84	6.864	-6.2	16
7awu	7.184	-5.8	20
5rgy	7.597	-6.9	22
5rgk	8.242	-6.4	18
5reh	8.626	-6.2	18

In Figure S1, we show the correlation plot of the RMSD's of the best binding poses on the 28 structures of Table S1 predicted using the full protomer (x-axis) or domain I+II alone (y-axis). The set includes 84 points corresponding to replicates of Vina calculations using box sizes of 20, 25 and 30 Å. The 4 outliers ($\Delta RMSD > 3$ Å) correspond to degenerate poses (i.e. having the same docking score).

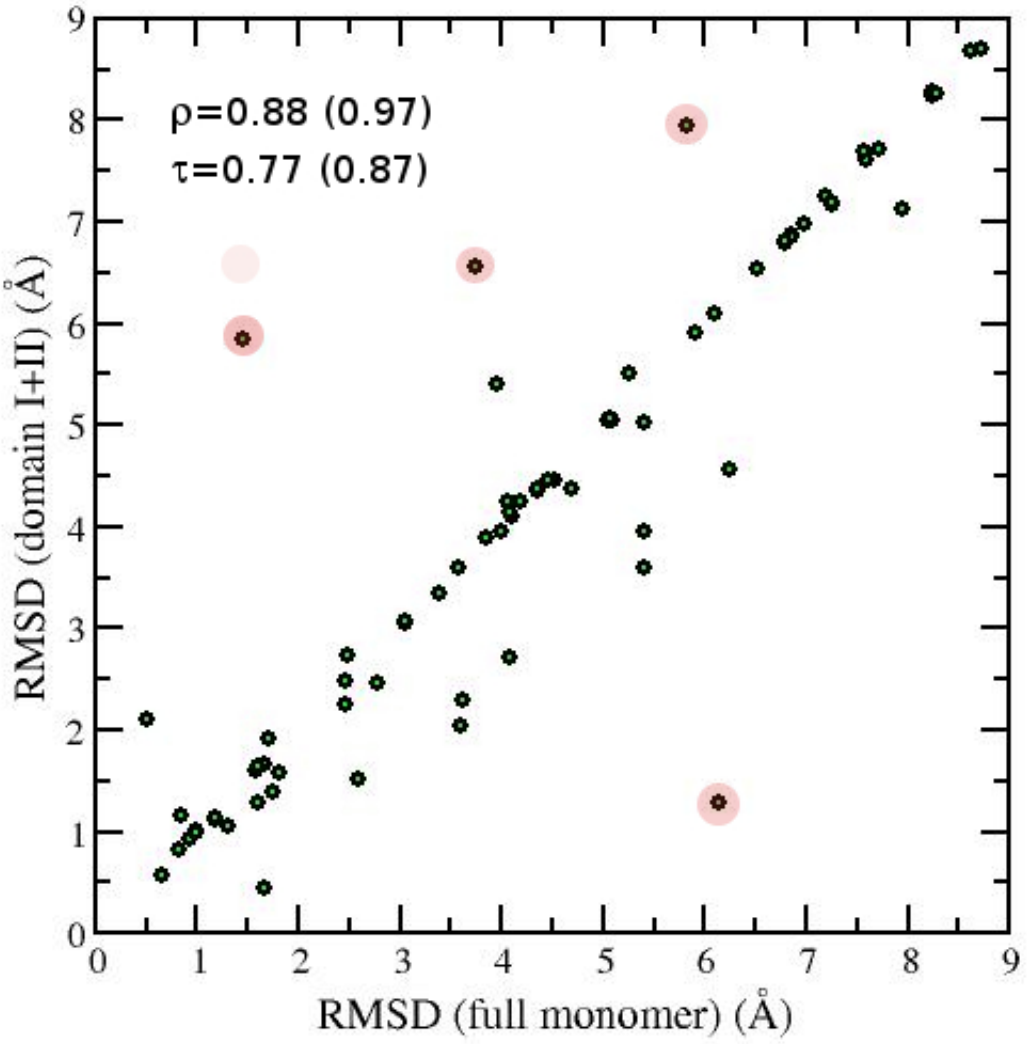
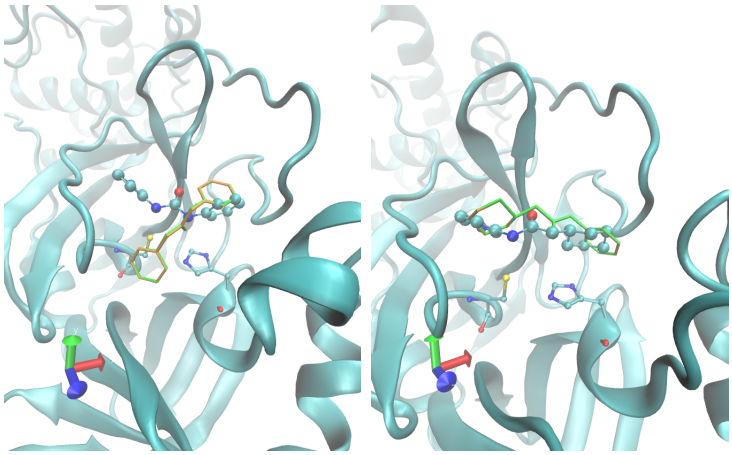
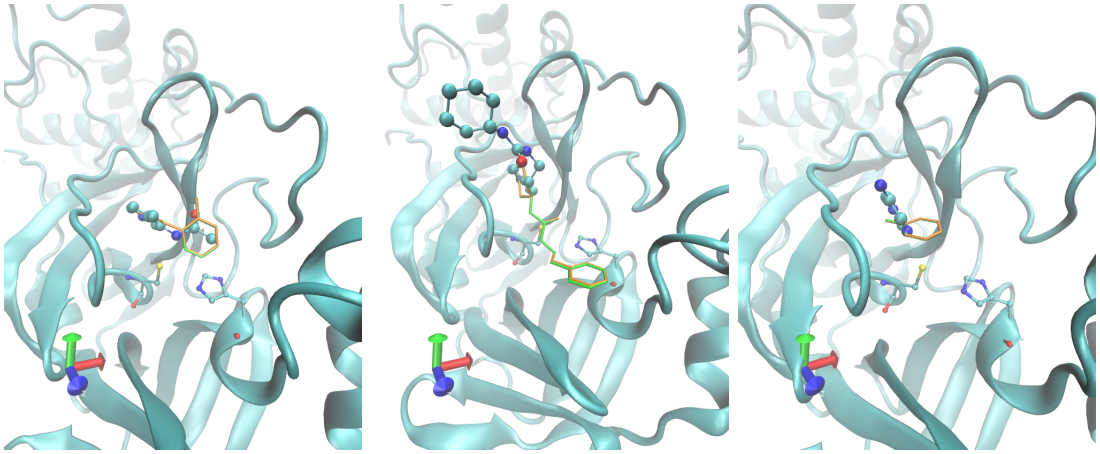


Figure S1: RMSD correlation plot for the full monomer and domain I+II. Outliers are marked in red. The value of the Pearson (ρ) and Kendall (τ) coefficients are reported. In parenthesis the corresponding values without the 4 outliers.

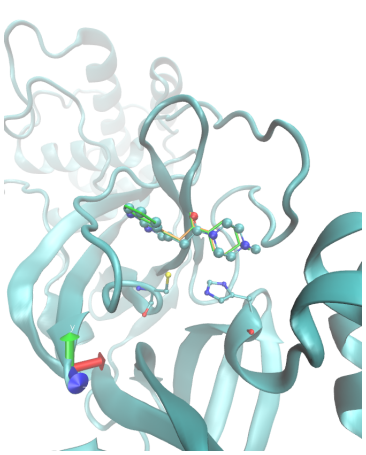
Figure S2: Structural comparison between best binding pose predicted using the full protomer (in orange) and the domain I+II (in green) for 5r83 (1), 5r84 (2), 5re4 (3), 5reh (4), 5rf3 (5), 5rf7 (6), 5rg1 (7), 5rgi (8), 5rgk (9), 5rgu (10), 5rgv (11), 5rgw (12), 5rgx (13), 5rgy (14), 5rgz (15), 5rh0 (16), 5rh1 (17), 5rh2 (18), 5rh3 (19), 5rh8 (20), 6m2n (21), 6w63 (22), 6w79 (23), 7aha (24), 7ans (25), 7awu (26), 7ju7 (27), 7kx5 (28). In cpk draw style it is reported the starting pose.



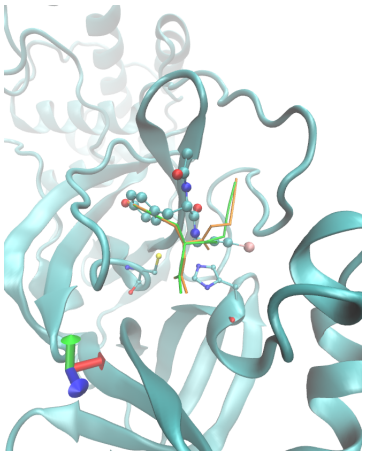
(1) **5r83** (2) **5r84**



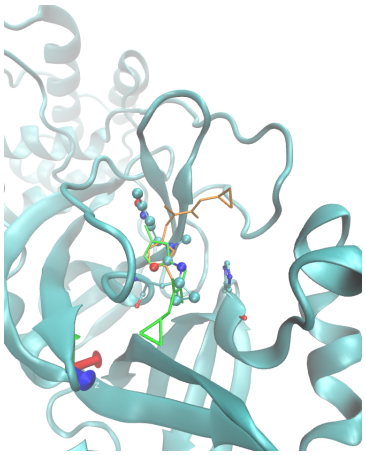
(3) **5re4** (4) **5reh** (5) **5rf3**



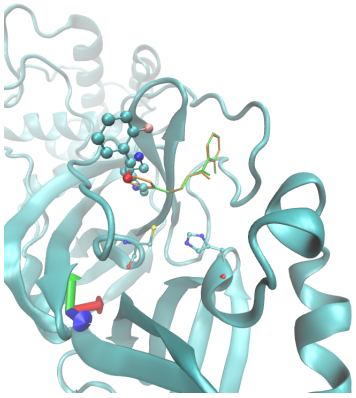
(6) **5rf7**



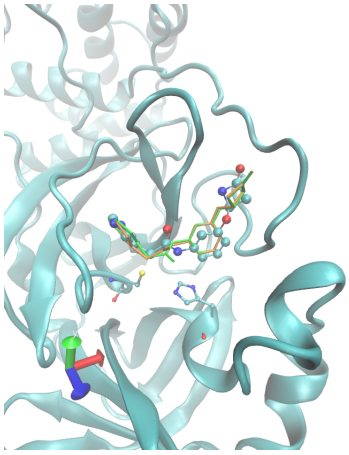
(7) **5rg1**



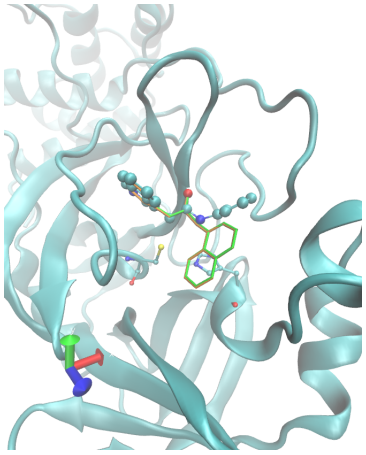
(8) **5rgi**



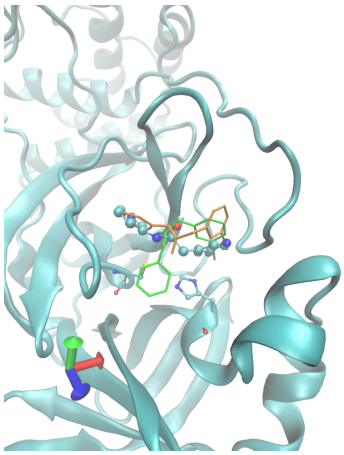
(9) **5rgk**



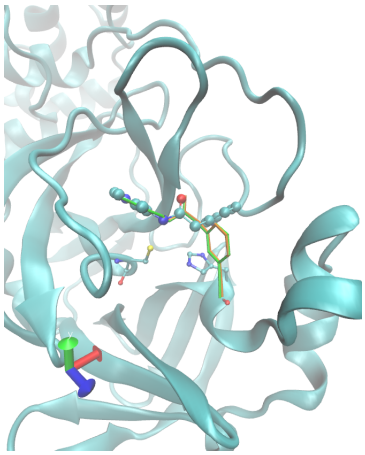
(10) **5rgu**



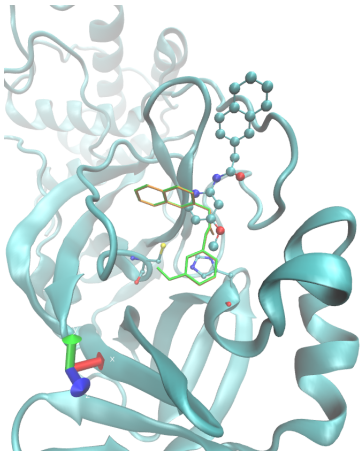
(11) **5rgv**



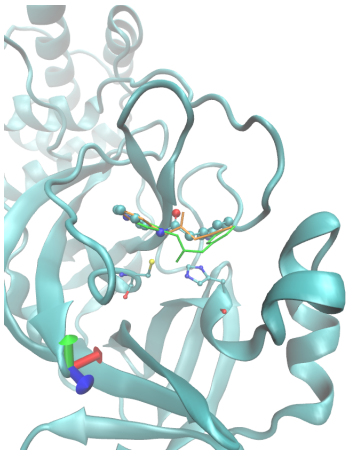
(12) **5rgw**



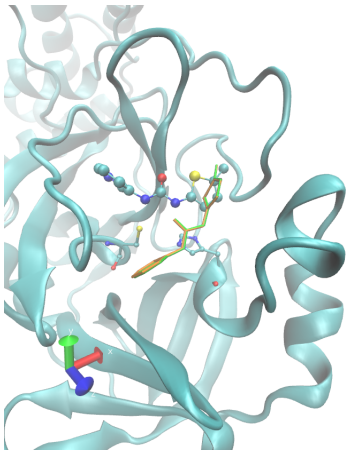
(13) **5rgx**



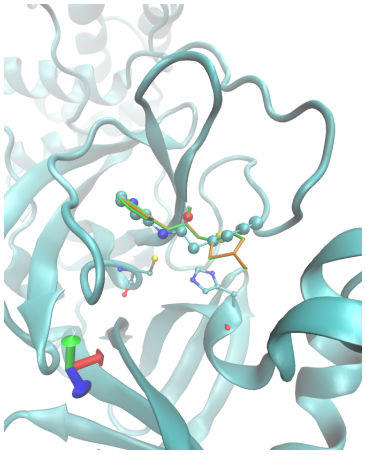
(14) **5rgy**



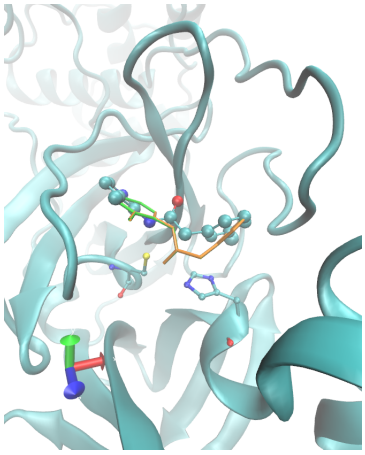
(15) **5rgz**



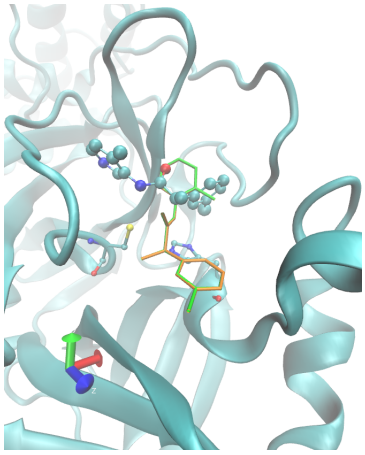
(16) **5rh0**



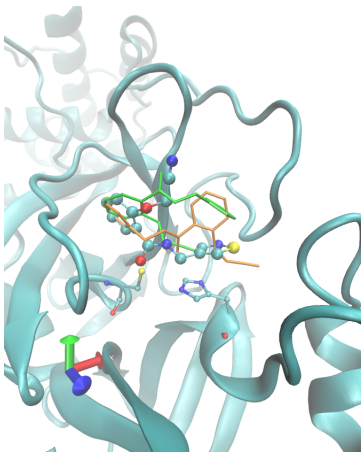
(17) **5rh1**



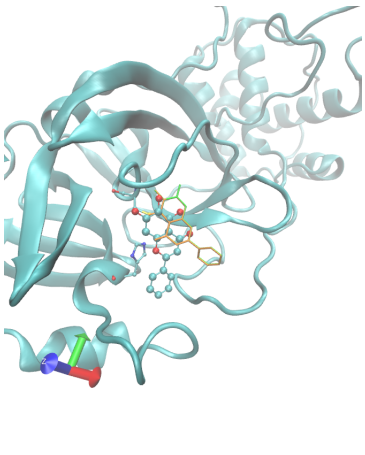
(18) **5rh2**



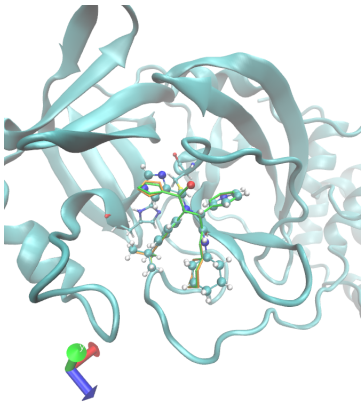
(19) **5rh3**



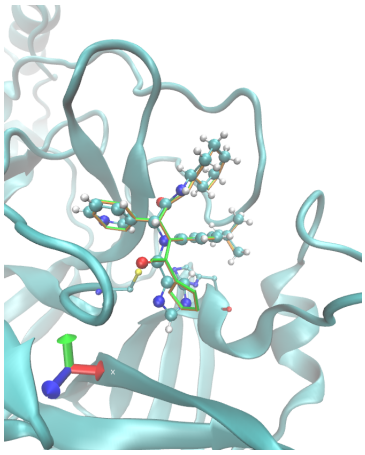
(20) **5rh8**



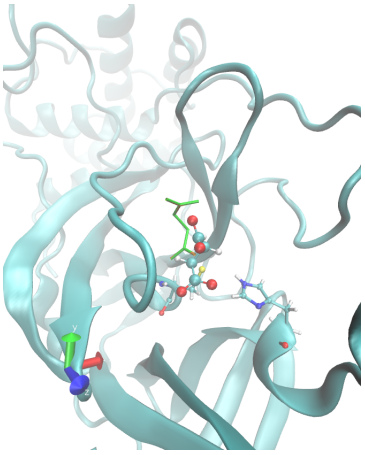
(21) **6m2n**



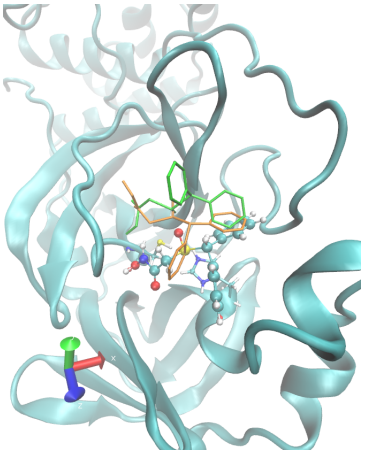
(22) **6w63**



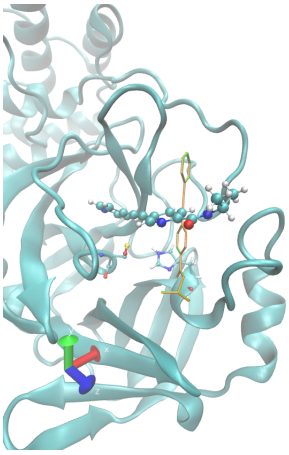
(23) **6w79**



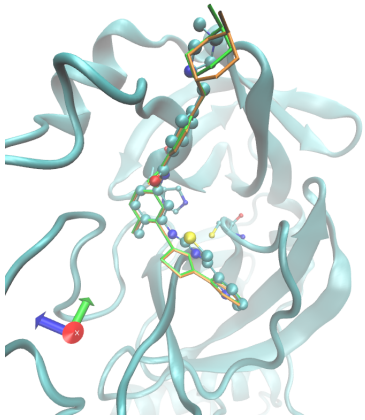
(24) **7aha**



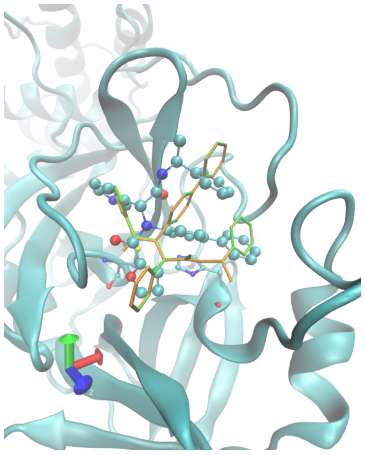
(25) **7ans**



(26) **7awu**



(27) **7ju7**



(28) **7kx5**

In Table S4, we report the Vina scoring functions and the corresponding distances between the PF-07321332 nitrile carbon and the S_γ of CYS145 in the 9 poses for both isoforms.

Table S4: Vina results on the H-C and H^+-C^- isoforms. ΔG and d_{S-C} are the binding free energy (in kcal/mol) and the distance (in Å) between the nitrile carbon in PF-07321332 and the S_γ of CYS145.

	H-C		H^+-C^-	
n_p	ΔG	d_{S-C}	ΔG	d_{S-C}
1	-7.9	8.2	-7.7	9.5
2	-7.5	5.7	-7.6	7.7
3	-7.3	4.9	-7.6	12.1
4	-7.1	10.0	-7.2	11.9
5	-6.9	4.6	-7.2	11.6
6	-6.8	6.5	-7.1	6.8
7	-6.6	4.3	-7.0	6.2
8	-6.6	4.4	-6.8	9.9
9	-6.6	8.6	-6.8	10.8

Molecular Dynamics Simulation

Starting from the best scoring ligand pose, the system was initially minimized at 0 K with a steepest descent procedure. The system was explicitly solvated with the TIP3P[??] water model at the standard density using GROMACS tools[??] and subsequently heated at 298.15 K in a the NPT ensemble (P=1 atm) using the Berendsen barostat[?] and a Bussi thermostat[?] with an integration time step of 0.1 fs and a coupling constant of 0.1 ps for 250 ps. Production runs in the NPT ensemble (T=298.15 and p=1 Atm) were carried out for a total of 100 ns imposing rigid constraints only on the X-H bonds (with X being any heavy atom) by means of the LINCS algorithm (δt=2.0 fs).[?] Electrostatic interactions were treated by using particle-mesh Ewald (PME)[?] method with a grid spacing of 1.2 Å and a spline interpolation of order 4. The total system charge (-1 e for both dyad states) was

neutralized by a uniform background plasma.[?] The cross interactions for Lennard-Jones terms were calculated using mixing rules in agreement with the FF used. The simulations and the trajectories analysis were performed using the GROMACS 2018.3 program.[? ?]

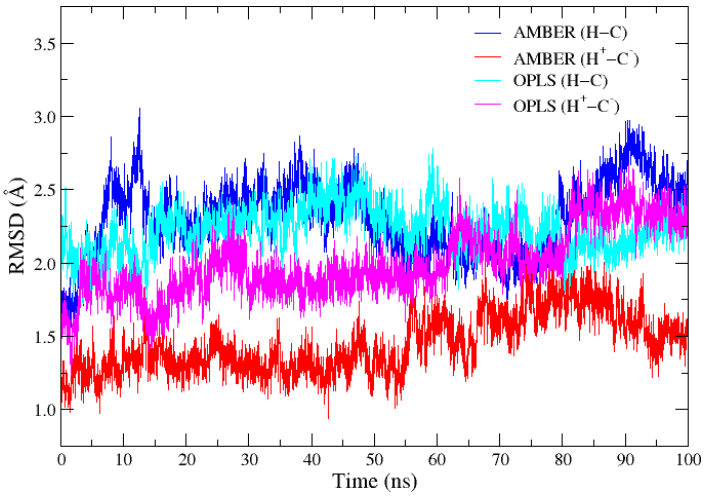


Figure S3: RMSD values of protein backbone calculated from MD simulations using the two FFs and the two states of the catalytic dyad (H41-C145). In blue AMBER FF and neutral dyad, in red AMBER FF and zwitterionic dyad, in cyan OPLS FF and neutral dyad and in magenta OPLS FF and zwitterionic dyad. The reference structure is the starting docking structure.

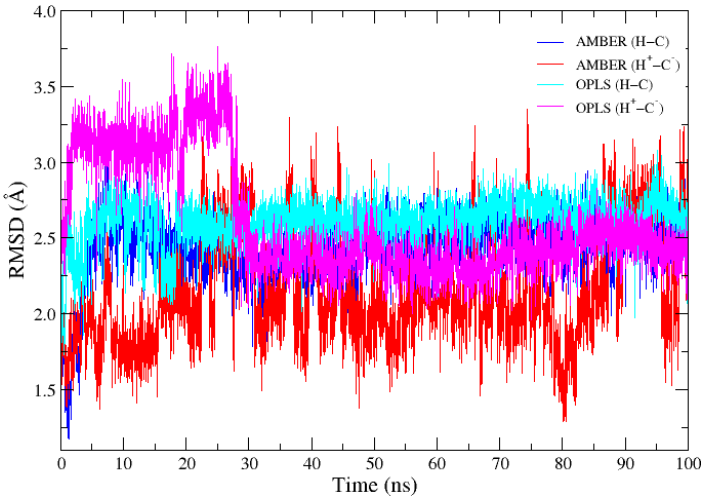


Figure S4: RMSD values of inhibitor calculated from MD simulations using the two FFs and the two states of the catalytic dyad (H41-C145). In blue AMBER FF and neutral dyad, in red AMBER FF and zwitterionic dyad, in cyan OPLS FF and neutral dyad and in magenta OPLS FF and zwitterionic dyad. The reference structure is the starting docking structure.

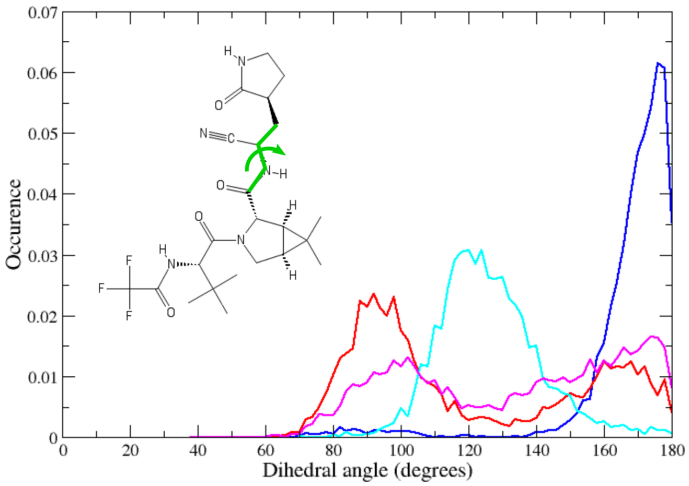


Figure S5: Histogram of dihedral angle showed in inset. In blue AMBER FF and neutral dyad, in red AMBER FF and zwitterionic dyad, in cyan OPLS FF and neutral dyad and in magenta OPLS FF and zwitterionic dyad.

References

() Zev, S.; Raz, K.; Schwartz, R.; Tarabeh, R.; Gupta, P. K.; Major, D. T. *J. Chem. Inf. Model.* **2021**, *61*, 2957–2966.

() O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminf.* **2011**, *3*, 33.

() Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

() Pagliai, M.; Macchiagodena, M.; Procacci, P.; Cardini, G. *J. Phys. Chem. Lett* **2019**, *10*, 6414–6418.

() Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845.

() Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

() Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

() Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

() Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

() Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

() Darden, T.; Pearlman, D.; Pedersen, L. G. *J. Chem. Phys.* **1998**, *109*, 10921–10935.