SUPPLEMENTARY INFORMATION: NNP/MM: Accelerating molecular dynamics simulations with machine learning potentials and molecular mechanics

Raimondas Galvelis,*,† Alejandro Varela-Rial,‡,¶ Stefan Doerr,‡ Roberto Fino,† Peter Eastman,§ Thomas E. Markland,§ John D. Chodera,‖ and Gianni De Fabritiis*,‡,¶,⊥

†*Acellera Labs, C/ Doctor Trueta 183, 08005 Barcelona, Spain*

‡*Acellera Ltd, Devonshire House 582, HA7 1JS, United Kingdom*

¶*Computational Science Laboratory, Universitat Pompeu Fabra, PRBB, C/ Doctor Aiguader 88, 08003 Barcelona, Spain*

§*Department of Chemistry, Stanford University, 337 Campus Drive, Stanford, CA, 94305, USA*

‖*Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA*

⊥*Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Spain*

E-mail: r.galvelis@acellera.com; g.defabritiis@gmail.com

Preparation and simulation conditions for the fragment system

The system preparation has been carried out with HTMD.[?] The fragment is constructed from SMILES and solvated with 5 Å padding.

The molecular topology and FF parameters files have been prepared with HTMD.[?] The force field parameters are obtained with GAFF2[?] or modeled with ANI-2x[?] depending on the simulation method. For water, the rigid TIP3P[?] model is used. The bonds between hydrogens and heavy atoms (H−X) are constrained and the hydrogen masses are repartitioned to 4.0 u. The electrostatic interactions are computed with PME[?] and the van der Waals interactions with the cutoff of 9.0 Å. The MD time step for the MM simulations is 4.0 fs, but for NNP/MM it is reduced to 2.0 fs due to stability issues. The temperature and pressures are maintained with the Langevin thermostat (the damping constant of 0.1 ps^{-1}) and the isotropic Monte Carlo barostat, respectively.

For metadynamics (MTD), PLUMED[?] plugin is used. Two dihedral angles are used as collective variables. The bias potential is updated every 1 ps, kernel height is 4.185 kcal/mol and width 10 deg. The bias factor is set to 5 (i.e. the effective temperature is 1550 K).[?]

Each protein-ligand complex has been equilibrated with the following procedure. First, the system is minimized for 500 steps. Second, the system is equilibrated for 1.0 ns with the NPT ensemble (T = 310 K and P = 1.0 bar). For the production simulations, the NVT ensemble (T = 310 K) is used without any restraints. Simulation trajectories are saved every 50 ps.

Dihedral angle series of the fragment

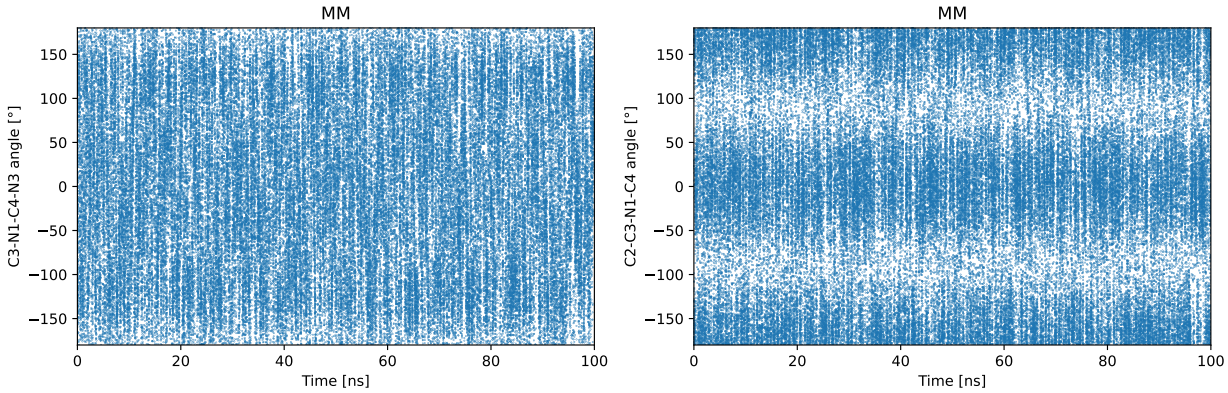

Figure S1: The dihedral angle series of the fragment from the MM simulation.

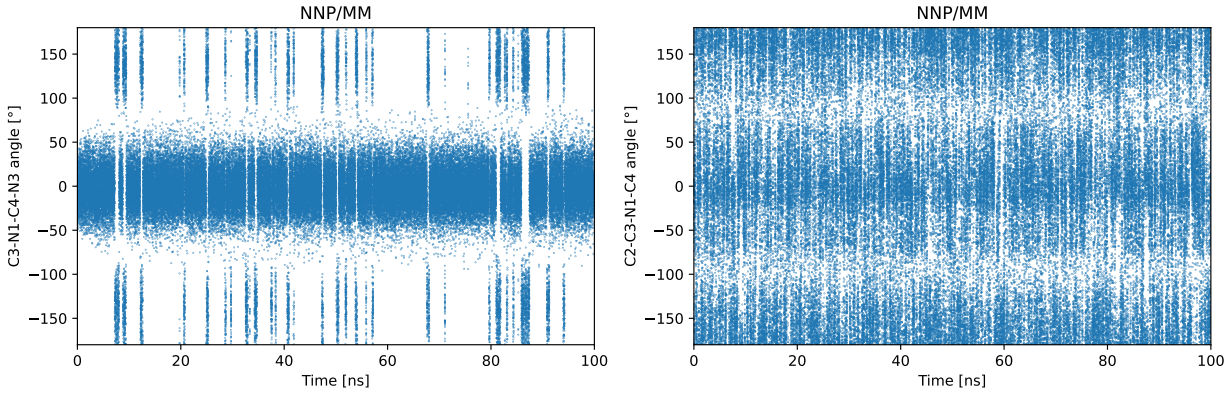

Figure S2: The dihedral angle series of the fragment from the NNP/MM simulation.

Preparation and simulation conditions for protein-ligand systems

The system preparation has been carried out with HTMD.[?] First, the structure of a protein-ligand complex is downloaded from the PDB[? ?] database. Only the protein and ligand are kept, and the other atoms are removed. Second, the protein is protonated at pH = 7.0 with PROPKA[? ?] and its hydrogen bonds are optimized with PDB2PQR.[? ?] Finally, the system is solvated with 5 Å padding and neutralized with Na^+ or Cl^- ions.

The molecular topology and FF parameters files have been prepared with HTMD.[?] For the protein, AMBER ff14SB[?] FF is used. For the ligand, the force field parameters are obtained with GAFF2[?] or modeled with ANI-2x[?] depending on the simulation method. For water, the rigid TIP3P[?] model is used. The bonds between hydrogens and heavy atoms (H−X) are constrained and the hydrogen masses are repartitioned to 4.0 u. The electrostatic interactions are computed with PME[?] and the van der Waals interactions with the cutoff of 9.0 Å. The MD time step for the MM simulations is 4.0 fs, but for NNP/MM it is reduced to 2.0 fs due to stability issues. The temperature and pressures are maintained with the Langevin thermostat (the damping constant of 0.1 ps^{-1}) and the isotropic Monte Carlo barostat, respectively.

Each protein-ligand complex has been equilibrated with the following procedure. First, the system is minimized for 500 steps with the harmonic positional restraint applied to the protein (the force constant for backbone atoms is 1.0 kcal/mol/$Å^2$ and the one for the side-chain atoms is 1.0 kcal/mol/$Å^2$). Second, the system is equilibrated for 1.0 ns with the NPT ensemble (T = 310 K and P = 1.0 bar). The force constants of the restraints are reduced linearly to 1.0 kcal/mol/$Å^2$ during the first 0.5 ns of the simulation. For the production simulations, the NVT ensemble (T = 310 K) is used without any restraints. Simulation trajectories are saved every 50 ps.

Dihedral angle scans

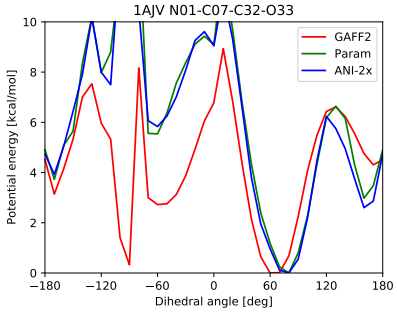

Figure S3: Energy profile of the dihedral angle N01-C07-C32-O33 scan of 1AJV ligand. Rotamer energies are computed with different methods.

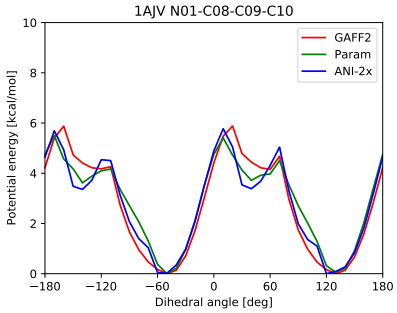

Figure S4: Energy profile of the dihedral angle N01-C08-C09-C10 scan of 1AJV ligand. Rotamer energies are computed with different methods.

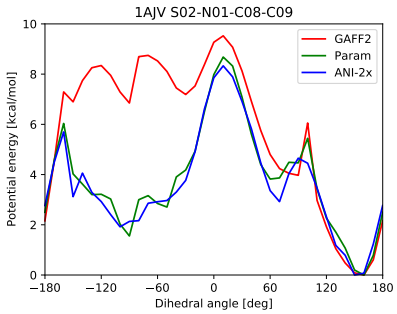

Figure S5: Energy profile of the dihedral angle S02-N01-C08-C09 scan of 1AJV ligand. Rotamer energies are computed with different methods.

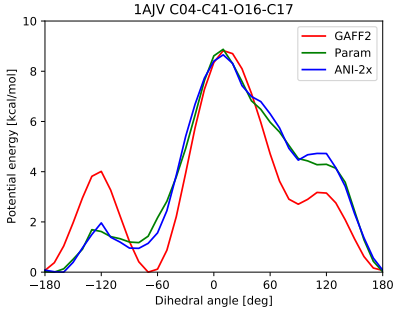

Figure S6: Energy profile of the dihedral angle C04-C41-O16-C17 scan of 1AJV ligand. Rotamer energies are computed with different methods.

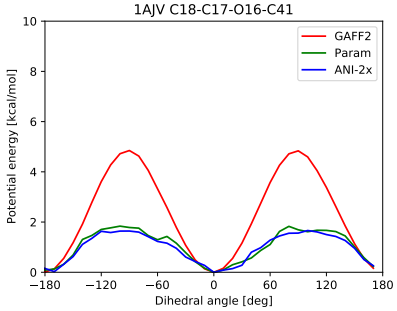

Figure S7: Energy profile of the dihedral angle C18-C17-O16-C41 scan of 1AJV ligand. Rotamer energies are computed with different methods.

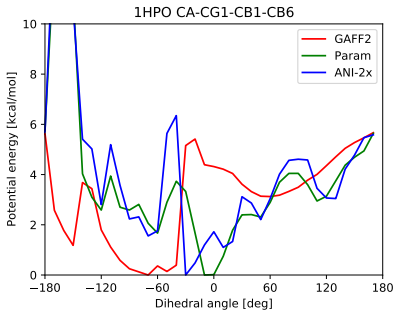

Figure S8: Energy profile of the dihedral angle CA-CG1-CB1-CB6 scan of 1HPO ligand. Rotamer energies are computed with different methods.

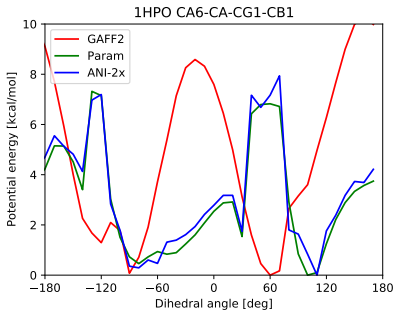

Figure S9: Energy profile of the dihedral angle CA6-CA-CG1-CB1 scan of 1HPO ligand. Rotamer energies are computed with different methods.

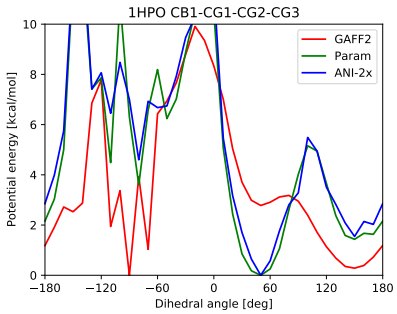

Figure S10: Energy profile of the dihedral angle CB1-CG1-CG2-CG3 scan of 1HPO ligand. Rotamer energies are computed with different methods.

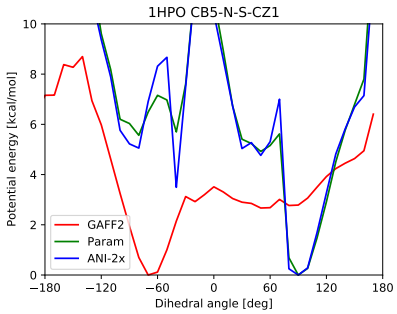

Figure S11: Energy profile of the dihedral angle CB5-N-S-CZ1 scan of 1HPO ligand. Rotamer energies are computed with different methods.

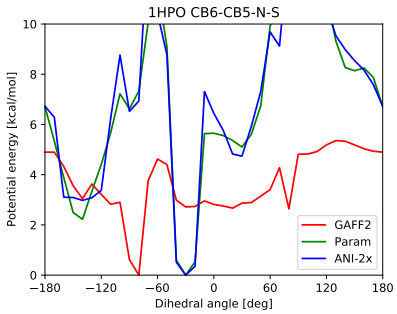

Figure S12: Energy profile of the dihedral angle CB6-CB5-N-S scan of 1HPO ligand. Rotamer energies are computed with different methods.

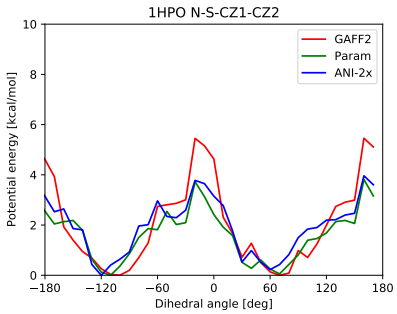

Figure S13: Energy profile of the dihedral angle N-S-CZ1-CZ2 scan of 1HPO ligand. Rotamer energies are computed with different methods.

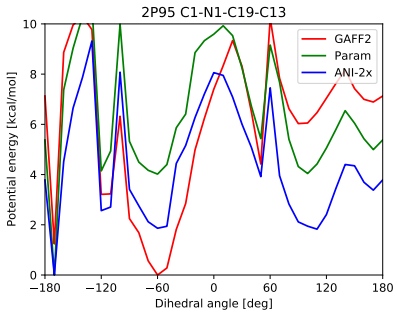

Figure S14: Energy profile of the dihedral angle C1-N1-C19-C13 scan of 2P95 ligand. Rotamer energies are computed with different methods.

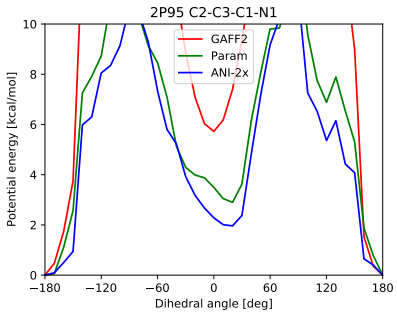

Figure S15: Energy profile of the dihedral angle C2-C3-C1-N1 scan of 2P95 ligand. Rotamer energies are computed with different methods.

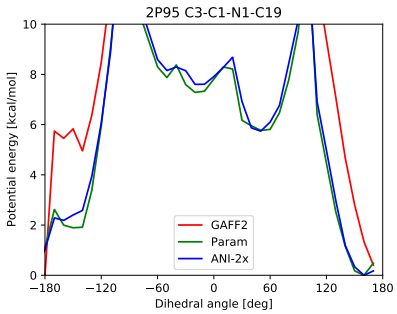

Figure S16: Energy profile of the dihedral angle C3-C1-N1-C19 scan of 2P95 ligand. Rotamer energies are computed with different methods.

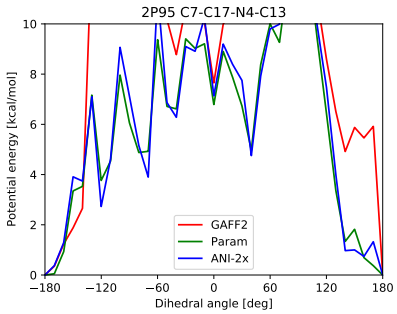

Figure S17: Energy profile of the dihedral angle C7-C17-N4-C13 scan of 2P95 ligand. Rotamer energies are computed with different methods.

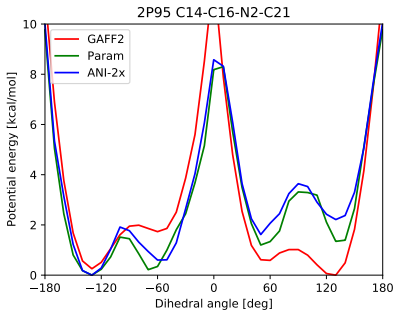

Figure S18: Energy profile of the dihedral angle C14-C16-N2-C21 scan of 2P95 ligand. Rotamer energies are computed with different methods.

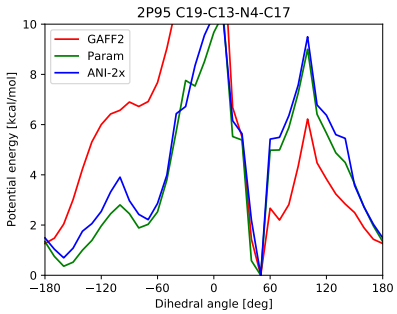

Figure S19: Energy profile of the dihedral angle C19-C13-N4-C17 scan of 2P95 ligand. Rotamer energies are computed with different methods.

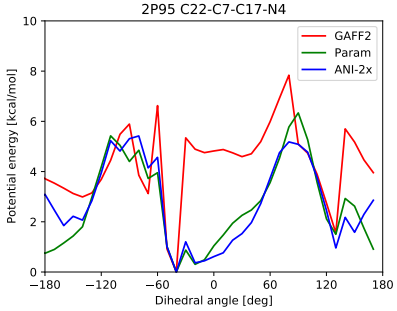

Figure S20: Energy profile of the dihedral angle C22-C7-C17-N4 scan of 2P95 ligand. Rotamer energies are computed with different methods.

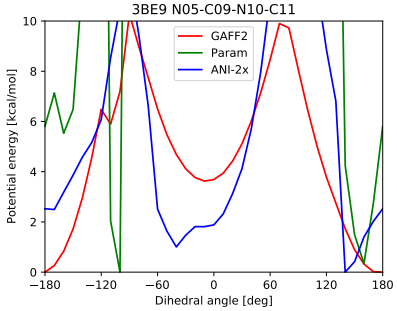

Figure S21: Energy profile of the dihedral angle N05-C09-N10-C11 scan of 3BE9 ligand. Rotamer energies are computed with different methods.

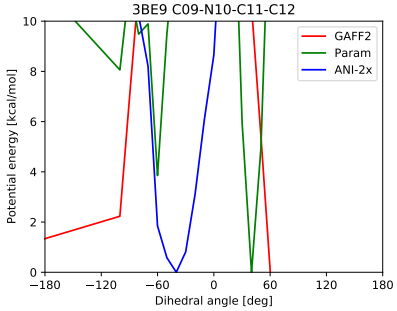

Figure S22: Energy profile of the dihedral angle C09-N10-C11-C12 scan of 3BE9 ligand. Rotamer energies are computed with different methods.

Time series of protein RMSD

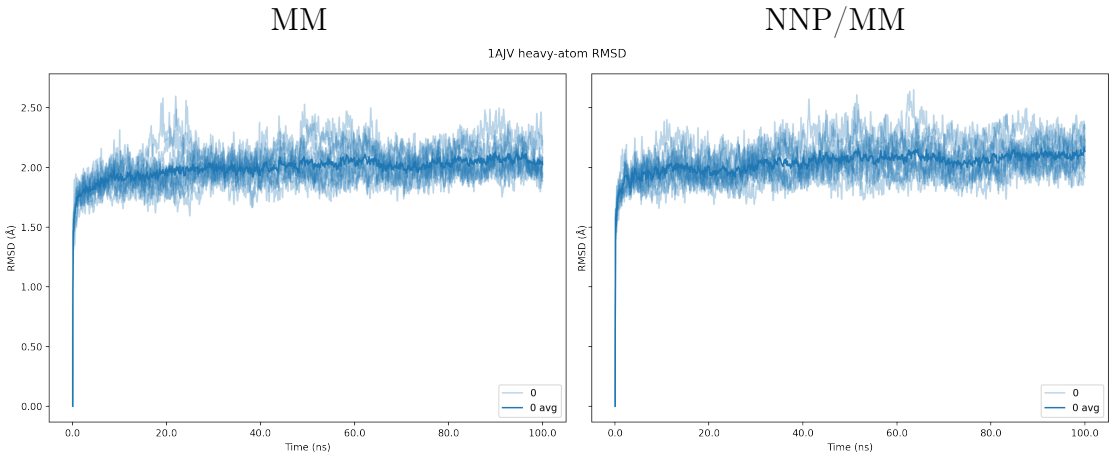

Figure S23: Time series of protein RMSD for 1AJV with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

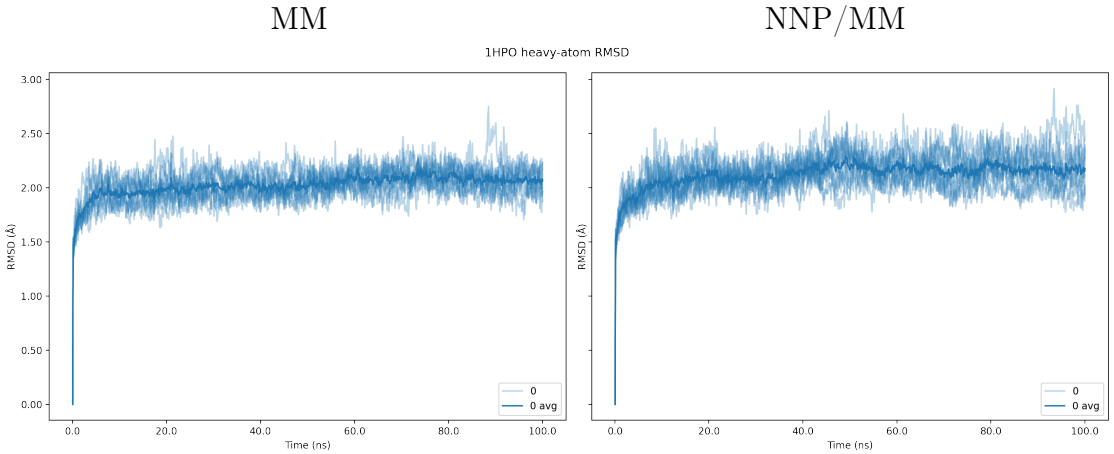

Figure S24: Time series of protein RMSD for 1HPO with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

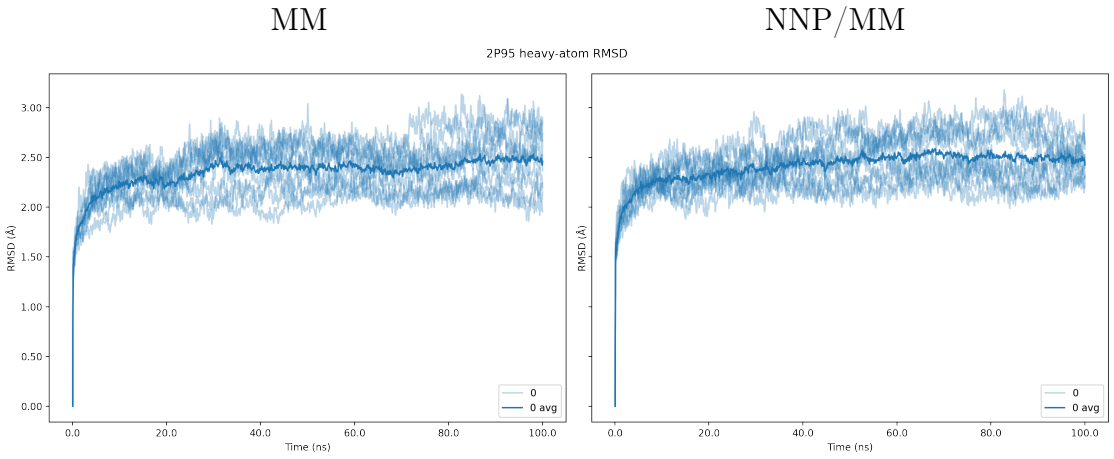

Figure S25: Time series of protein RMSD for 2P95 with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

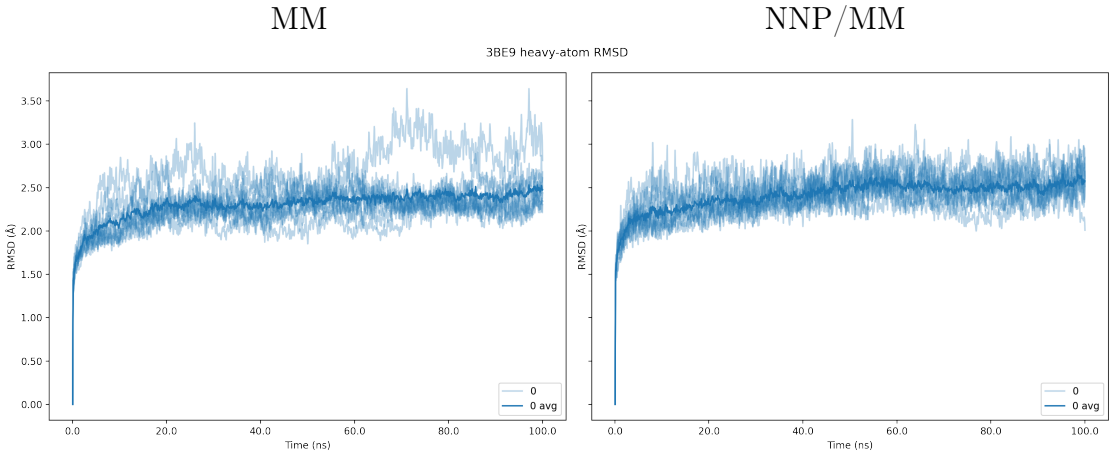

Figure S26: Time series of protein RMSD for 3BE9 with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

Per-residue RMSF of protein

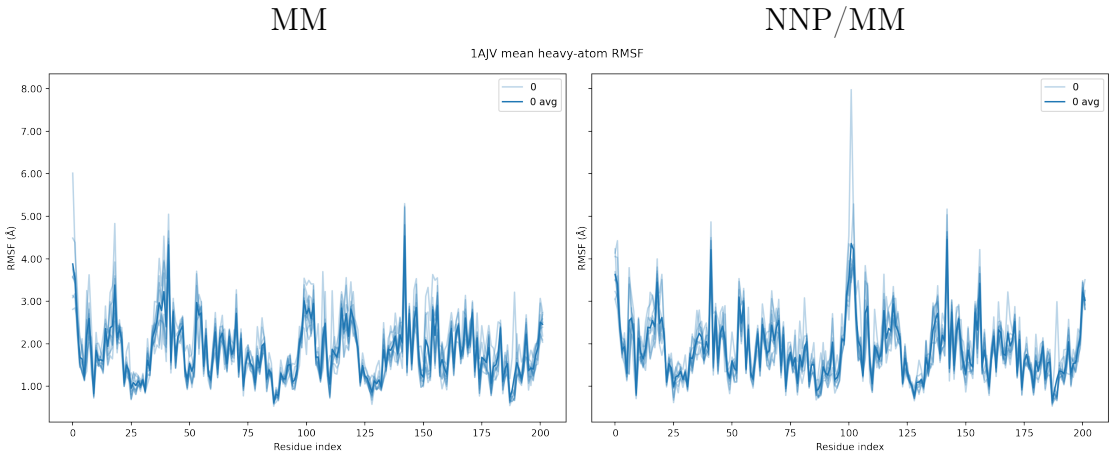

Figure S27: Residue RMSF of protein for 1AJV with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

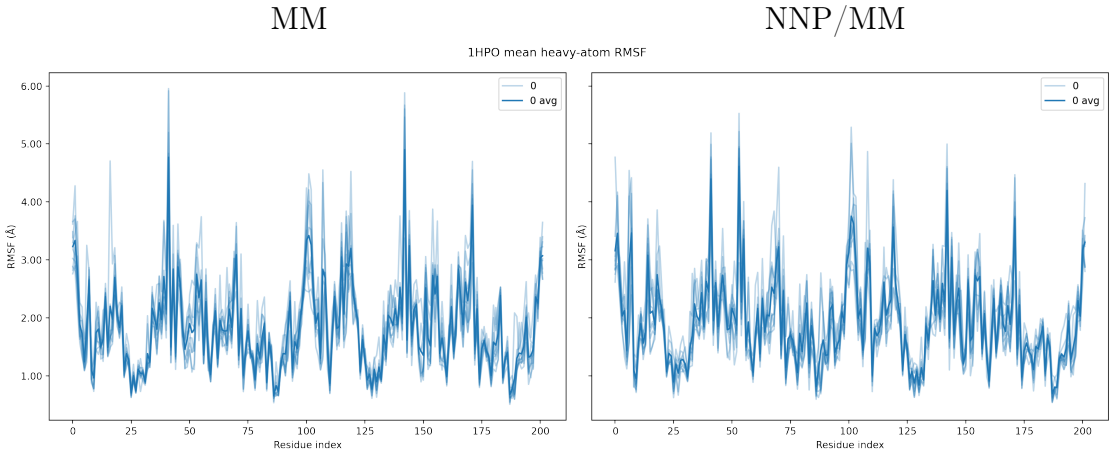

Figure S28: Residue RMSF of protein for 1HPO with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

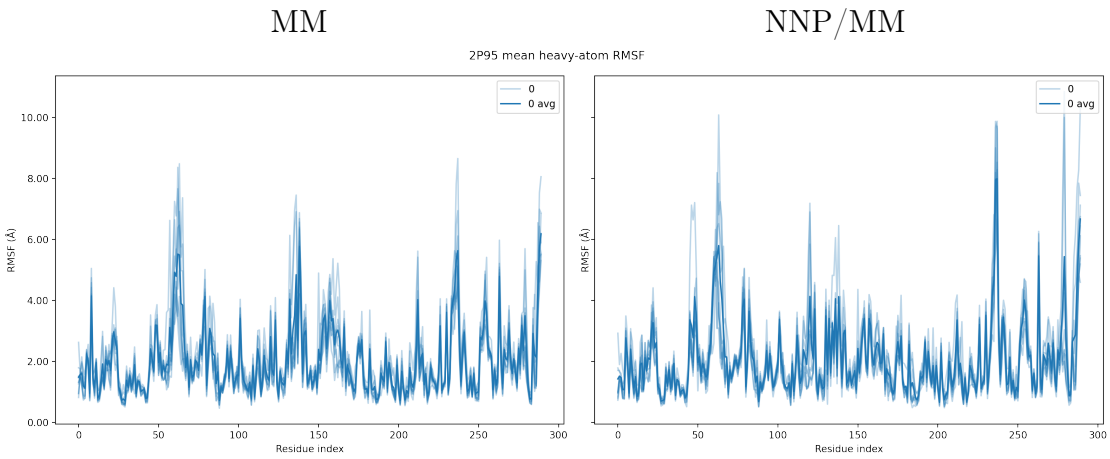

Figure S29: Residue RMSF of protein for 2P95 with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

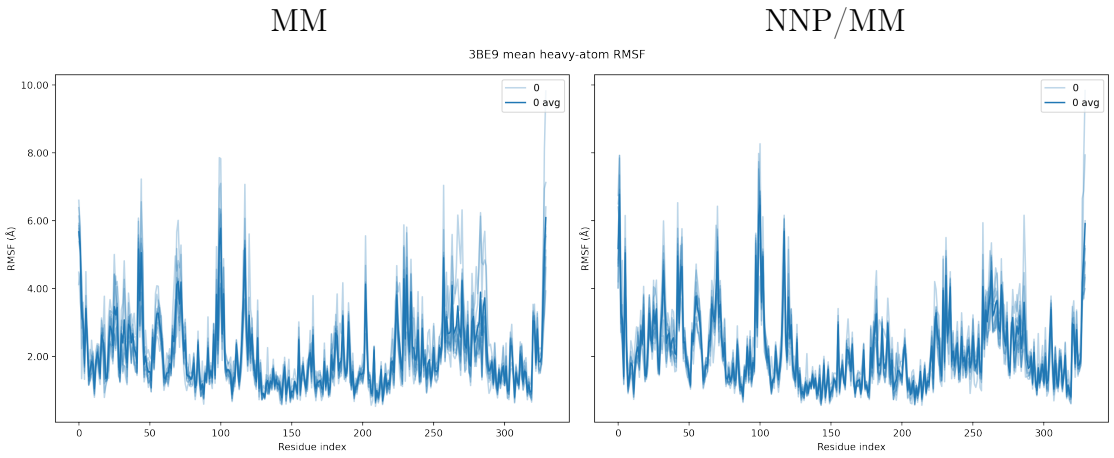

Figure S30: Residue RMSF of protein for 3BE9 with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

Time series of ligand RMSD

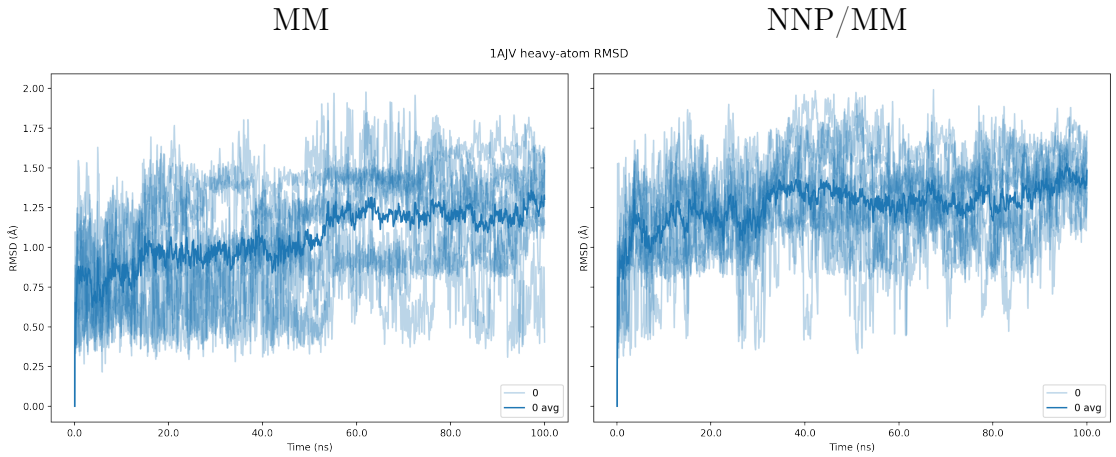

Figure S31: Time series of ligand RMSD for 1AJV with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

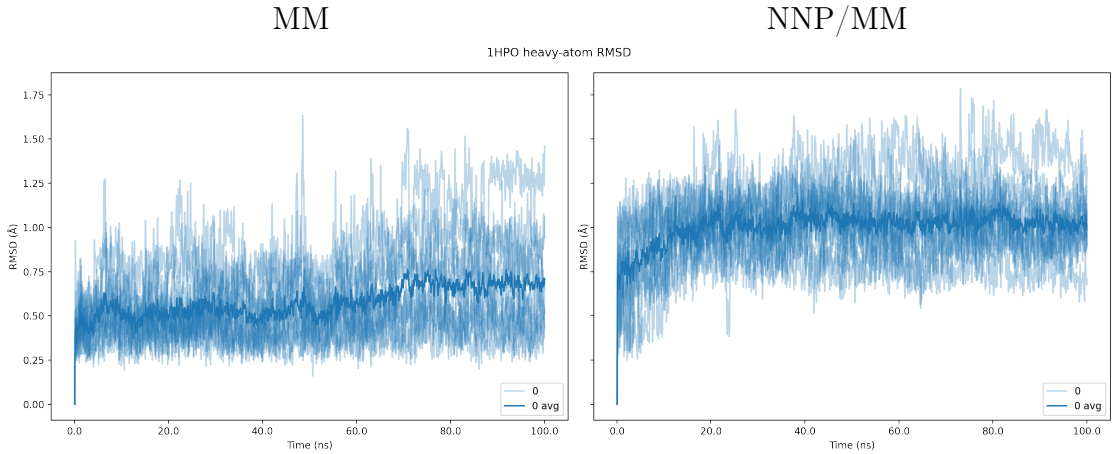

Figure S32: Time series of ligand RMSD for 1HPO with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

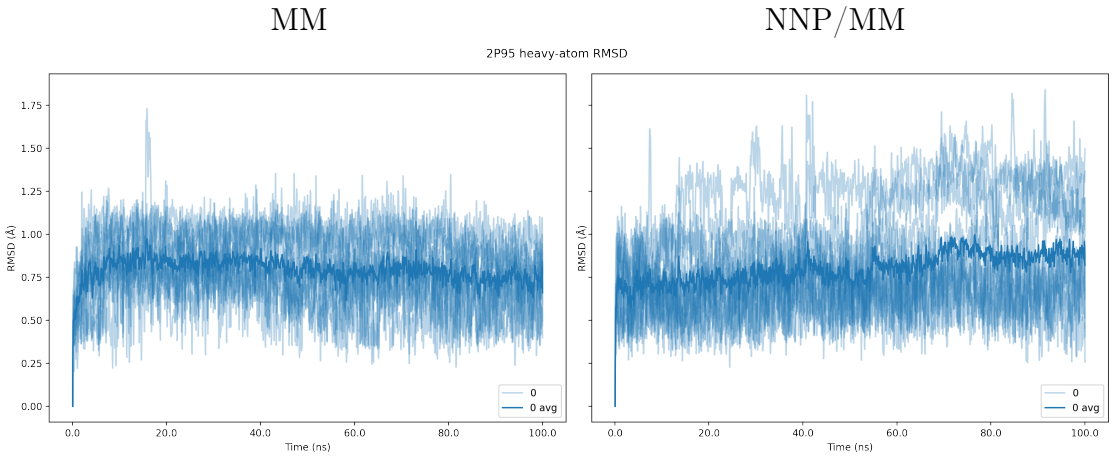

Figure S33: Time series of ligand RMSD for 2P95 with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

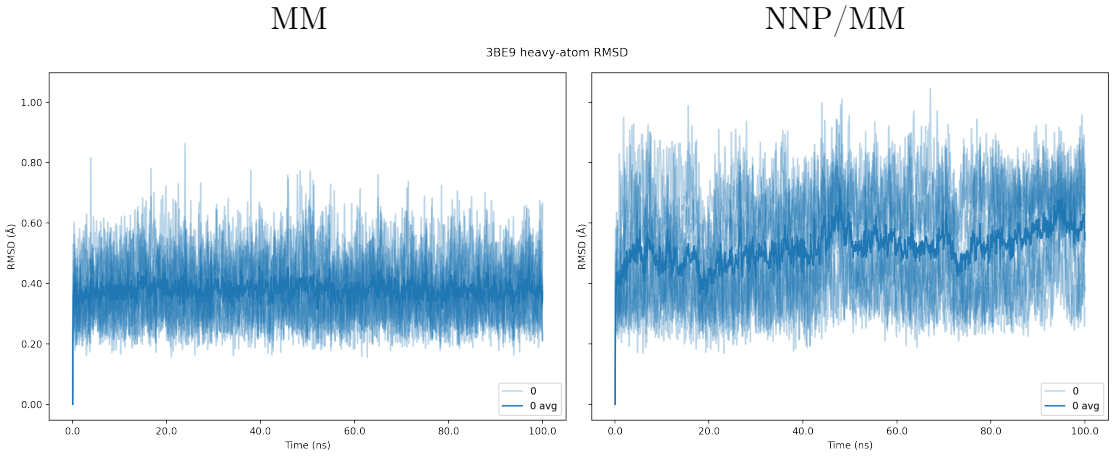

Figure S34: Time series of ligand RMSD for 3BE9 with the different methods (MM and NNP/MM). The independent simulations are colored in light blue and the average is in dark blue.

Protein-ligand interactions

Protein-ligand interactions have been detected following criteria:

- H-bond: the donor and acceptor distance (D−A) is <2.5 Å and the D−H−A angle is >120°.
- π-π: the distance between the centers of aromatic rings (R−R) and the angle between their norms (N) satisfying one of these criteria:
  - R−R is <4.4 Å and N is <30°.
  - R−R is <5.5 Å and N is >60°.
- Cation-π: the distance between a positively-charged atom (P) and the center of the aromatic ring (R) is <5.0 Å and the angle of P−R with respect to the plane of R is >60°.
- σ-hole: the distance between halogen atom (H) and the center of the aromatic ring (R) is <4.0 Å and the angle of H−R with respect to the plane of R is >60°.

The trajectories have been sampled every 50 ps and the probabilities computed aggregating data from all the 10 independent simulations.

Table S1: Probability of 1AJV protein-ligand interactions in the MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
H-bond	ASP:26	0.87665
H-bond	ASP:26	0.99235
H-bond	ILE:51	0.96835
H-bond	ASH:127	0.26010
H-bond	ASH:127	0.51815
H-bond	ILE:152	0.88880

Table S2: Probability of 1AJV protein-ligand interactions in the NNP/MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
H-bond	ASP:26	0.99210
H-bond	ASP:26	0.99140
H-bond	ILE:51	0.96035
H-bond	ASH:127	0.05240
H-bond	ASH:127	0.11475
H-bond	ILE:152	0.91555

Table S3: Probability of 1HPO protein-ligand interactions in the MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
Cation-π	ARG:9	0.43165
H-bond	ASH:26	0.11160
H-bond	ILE:51	0.36980
H-bond	ILE:51	0.02740
H-bond	ASP:127	0.94340
H-bond	ASP:131	0.28660
H-bond	GLY:150	0.92555
H-bond	GLY:150	0.57730
H-bond	GLY:151	0.01885
H-bond	ILE:152	0.90360

Table S4: Probability of 1HPO protein-ligand interactions in the NNP/MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
Cation-π	ARG:9	0.76190
H-bond	ASH:26	0.10260
H-bond	ILE:51	0.38165
H-bond	ILE:51	0.10210
H-bond	ASP:127	0.99580
H-bond	ASP:131	0.81165
H-bond	ASP:132	0.01970
H-bond	GLY:150	0.99445
H-bond	GLY:150	0.11820
H-bond	ILE:152	0.89975

Table S5: Probability of 2P95 protein-ligand interactions in the MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
π-π	TYR:86	0.62565
π-π	TYR:86	0.10760
π-π	PHE:163	0.03485
π-π	PHE:163	0.57840
π-π	TRP:206	0.56830
H-bond	GLY:207	0.90965
H-bond	GLY:207	0.05580
H-bond	GLY:209	0.35475
σ-hole	TYR:219	0.54095

Table S6: Probability of 2P95 protein-ligand interactions in the NNP/MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
π-π	TYR:86	0.63505
π-π	TYR:86	0.08905
π-π	PHE:163	0.06140
π-π	PHE:163	0.59525
π-π	TRP:206	0.01760
π-π	TRP:206	0.64465
H-bond	GLY:207	0.89720
H-bond	GLY:207	0.07225
H-bond	GLY:209	0.54680
σ-hole	TYR:219	0.38280

Table S7: Probability of 3BE9 protein-ligand interactions in the MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
H-bond	LYS:64	0.68070
H-bond	VAL:112	0.57070
H-bond	VAL:112	0.38385
π-π	HID:156	0.24210
H-bond	ASP:171	0.26770
H-bond	ASP:171	0.10400

Table S8: Probability of 3BE9 protein-ligand interactions in the NNP/MM simulations. The interactions with a lower probability than 0.01 are excluded.

Interaction	Residue	Probability
H-bond	VAL:41	0.04610
H-bond	LYS:64	0.78265
H-bond	VAL:112	0.65315
H-bond	VAL:112	0.98575
π-π	HID:156	0.24225
H-bond	ASP:171	0.10545
H-bond	ASP:171	0.03135