Supporting Information: Protein-ligand free energies of binding from full-protein DFT calculations: convergence and choice of exchange-correlation functional

Lennart Gundelach, Thomas Fox, Christofer S. Tautermann, Chris-Kriton Skylaris*

March 2021

1 Further Computational Details

1.1 Input and Output files: https://github.com/gundelach/ESI-PCCP

All original input files for the DFT calculations in ONETEP, dispersion calculations in dftd3, normal mode calculations in AMBER and GFN2-XTB calculations in XTB are provided in the public GitHub repository:

https://github.com/gundelach/ESI-PCCP

Raw output files are also provided where appropriate. For details of the files provided, please consult the README documentation included with the files.

1.2 ONETEP Background

While traditional density functional theory scales cubically with the number of electrons in the system, linearly-scaling versions of DFT have been developed. The ONETEP, code [1] is one such linear-scaling DFT implementation. To achieve linear-scaling, the one-particle electron density matrix, which decays exponentially with distance, is truncated. Effectively, the electronic density is localized in space. In ONETEP this is achieved by representing localized orbitals by non-orthogonal generalized Warnier functions (NGWFs). As part of the self-consistent minimization, both the non-local orbitals as well as the electron density are optimized. The parallelization strategy of ONETEP exploits this locality and can spread individual atoms over many simultaneous processes. A hybrid MPI-OMP parallelization allows for efficient and scalable distribution of the workload. MPI processes split up the system into individual or groups of atoms and which spawn multiple OMP threads for each required computation.

The ONETEP implicit solvation model is a minimal-parameter Poisson-Boltzmann (PB) based model which is implemented self-consistently as part

of the DFT calculation[2]. The solute cavity is constructed from iso-surfaces of the electronic-density with a smooth transition between the bulk dielectric constant and the interior of the solute. Using the electronic density to construct the cavity reduces the number of empirical parameters. Because the PB equation is solved self-consistently, the implicit solvent can alter the electron density. The smeared-ion formalism is used to model nuclear cores and is controlled by a single parameter. The non-polar term of the solvation energy is calculated from the surface area of the iso-surfaces used to construct the cavity. The solvent accessible surface area (SASA) is multiplied by the physical surface tension of the solvent, γ, and scaled by an empirical factor of 0.28 to approximately include dispersion-repulsion between the solvent and solute

1.3 Cavity-correction

Both the minimal-parameter PBSA solvent-model implemented in ONETEP, as well as the PBSA solvent used at MM-level, incorrectly handle the buried cavity in T4-lysozyme (L99A/M102Q). This is a known issue for solvent models based on the solvent accessible surface area described in detail in 2010 by Genheden et al. [3] and in 2014 by Fox et al. [4]. In the un-complexed protein calculation, i.e the host, the surface area of the interior of the buried binding site is counted towards the solvent accessible surface area (SASA) used to calculate the non-polar solvation term. Thus, the non-polar term of just the protein is larger than that of the complex indicating the formation of a larger cavity in the solvent. Conceptually, the SASA model creates an additional, fictitious, cavity in the solvent with the SASA of the buried binding site. Because the non-polar term of both the protein and complex are known, a post-hoc cavity-correction may be applied to remove the additional contribution of the buried cavity to the non-polar solvation energy. By subtracting the difference in the non-polar terms of the host and complex from the host solvation energy, the spurious cavitation energy is removed. If the cavity should be filled with water, as is the case for this T4-lysozyme mutant (L99A/M102Q) [5], this correction is applied a second time to remove the cavity SASA from the host SASA. This is necessary because a water filled cavity reduces the volume of the cavity formed for the solute in the solvent.

This simple approach works immediately for the original MM-PBSA calculations, which used a simplistic SASA model with no estimate of the dispersion and repulsion between the solvent and solute (npopt=1 in Amber10). In the QM calculations, dispersion/repulsion is also included in the non-polar term. Because the cavity SASA does represent a physical boundary between solvent and solute inside the buried cavity, the dispersion-repulsion due to the cavity SASA term is physically correct. Since the non-polar dispersion in ONETEP is just a scaling factor applied to the non-polar term, it can be separated out and only the excess cavitation energy removed from the host non-polar solvation term. A full derivation is provided in [4].

1.4 ONETEP Settings

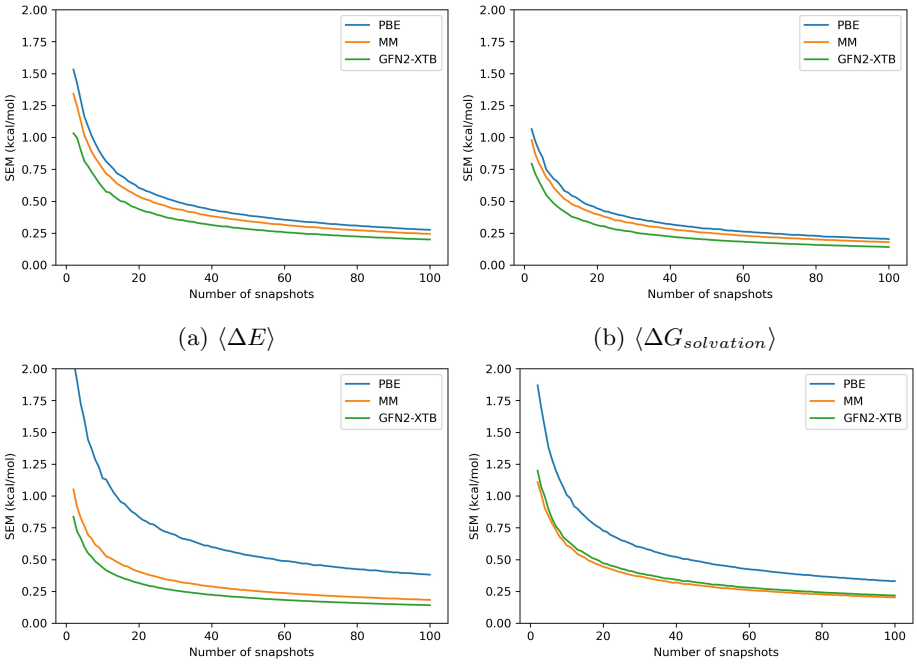
A kinetic energy cutoff of 800eV was used for all functionals. 4 NGWFs were used for Carbon, Nitrogen and Oxygen and 1 NGWF was used for Hydrogen. For Sulfur and Fluorine 9 NGWFS were used. An NGWF radius of $8a_0$ was used throughout. The Pseudoatomic Solver was used in ONETEP to generate the initial NGWFs. ONETEP default parameters for water at room temperature were used. The default solvent surface tension is 4.7624×10^{-5}Ha/Bohr^2 with an apolar scaling factor of 0.281075 and a solvation β of 1.3. The bulk permitivity was 78.54 and an interior dielectric of 1 was used. To speed up the solution of the Poisson-Boltzmann equation, the charge at the boundary of the simulation cell was coarse-grained. The default charge coarse-graining factor at the boundary is 5, but by increasing this to 10, the energy evaluation of the complex is sped up by 20%. In a series of test calculations, increasing the coarse-graining to 10 resulted in a change of only 0.005 kcal/mol in the total energy in solvent and 0.01 kcal/mol in the binding energies. ONETEP version 5.3.2.6 compiled with the Intel 2019 compilers and Intel MPI 2019 was used.

2 Data and Figures

2.1 Standard Error of the Mean

2.1.1 Enthalpy, Ligand Set B; 100 snapshots; PBE, MM, GFN2-XTB

Figures 1-7: The standard error of the mean (SEM) calculated by bootstrapping (1000 re-samples) of the change upon binding in the gas-phase energy, $\langle\Delta E\rangle$, solvation energy, $\langle\Delta G_{solvation}\rangle$, cavity-corrected solvation energy, $\langle\Delta G_{solvation-cav-cor}\rangle$, and total enthalpy, $\langle\Delta H_{bind}\rangle = \langle\Delta E\rangle + \langle\Delta G_{solvation}\rangle$, for each ligand up to 100 snapshots for MM, the DFT functional PBE and the SEQM method GFN2-XTB.



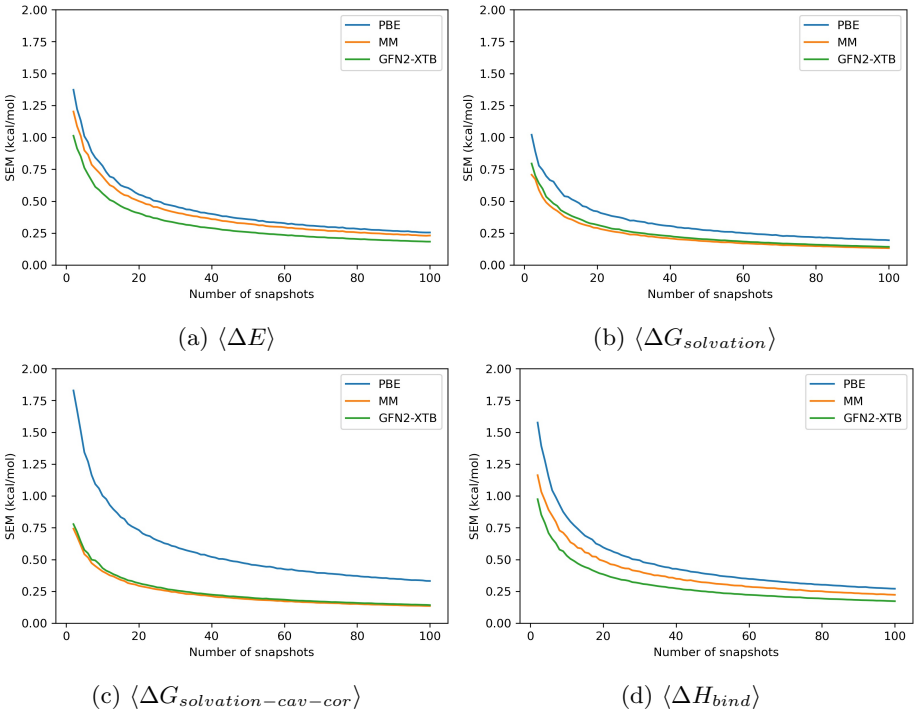
(a) $\langle \Delta E \rangle$

(b) $\langle \Delta G_{solvation} \rangle$

(c) $\langle \Delta G_{solvation-cav-cor} \rangle$

(d) $\langle \Delta H_{bind} \rangle$

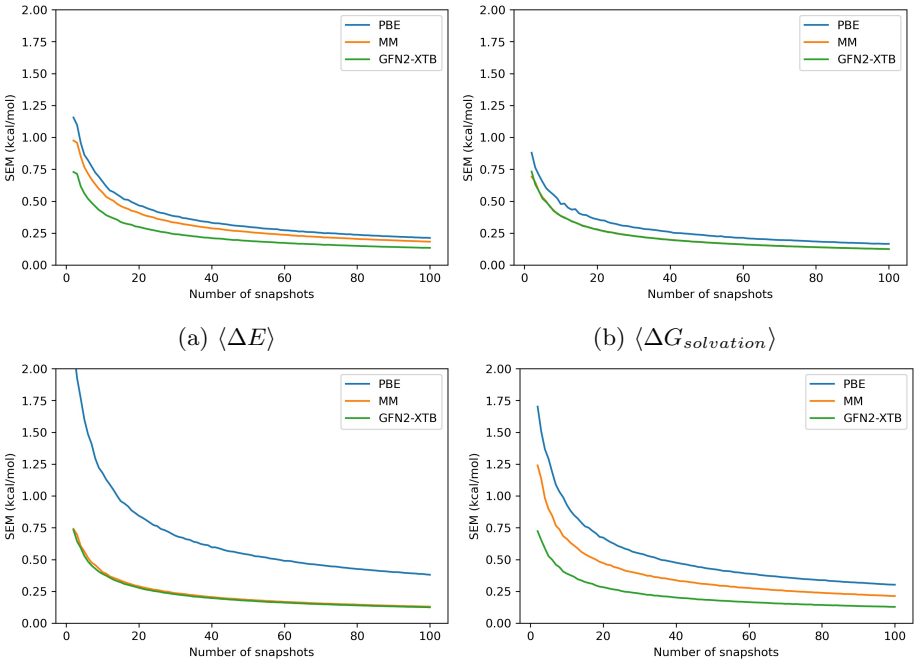
Figure 1: catechol



(a) $\langle \Delta E \rangle$ (b) $\langle \Delta G_{solvation} \rangle$

(c) $\langle \Delta G_{solvation-cav-cor} \rangle$ (d) $\langle \Delta H_{bind} \rangle$

Figure 2: phenol



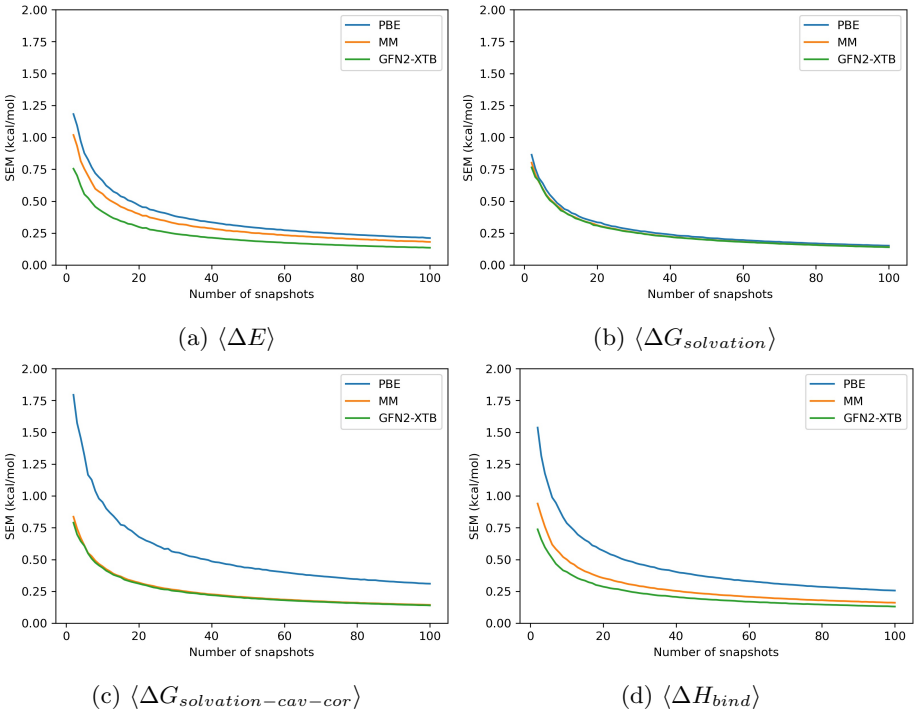
(a) $\langle \Delta E \rangle$

(b) $\langle \Delta G_{solvation} \rangle$

(c) $\langle \Delta G_{solvation-cav-cor} \rangle$

(d) $\langle \Delta H_{bind} \rangle$

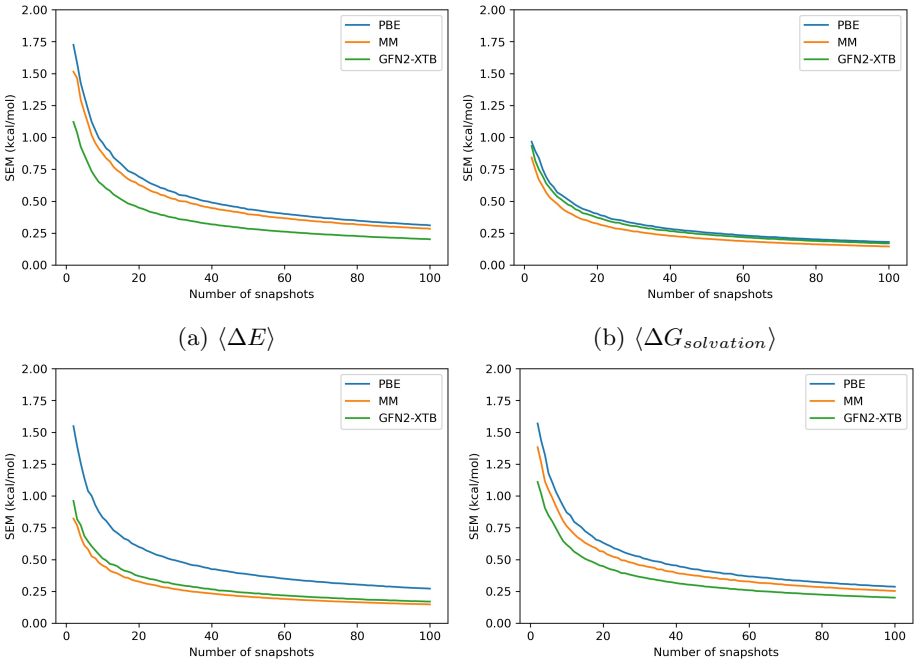
Figure 3: fluoroaniline



(a) $\langle \Delta E \rangle$ (b) $\langle \Delta G_{solvation} \rangle$

(c) $\langle \Delta G_{solvation-cav-cor} \rangle$ (d) $\langle \Delta H_{bind} \rangle$

Figure 4: methylphenol



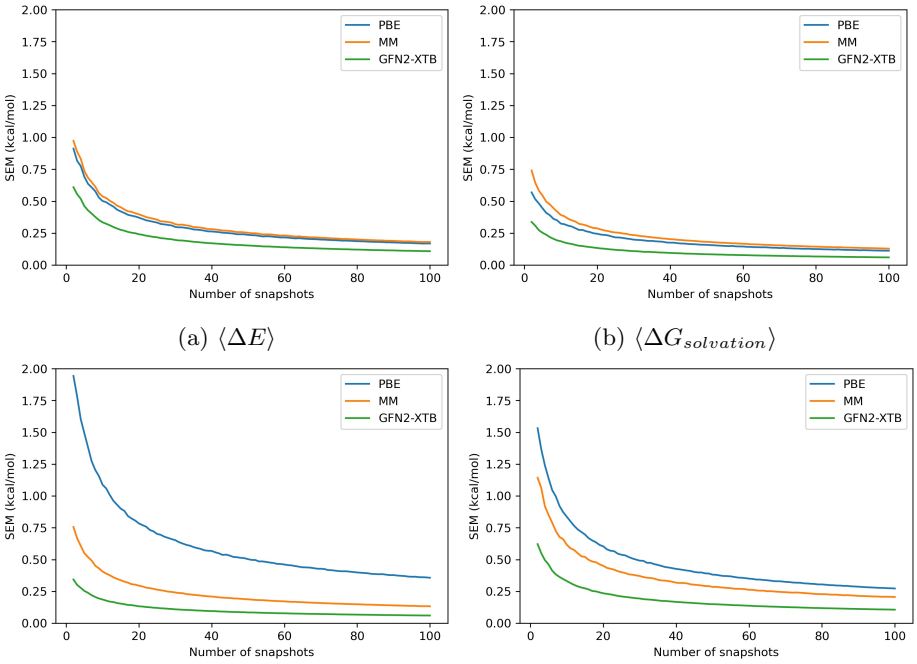
(a) $\langle \Delta E \rangle$

(b) $\langle \Delta G_{solvation} \rangle$

(c) $\langle \Delta G_{solvation-cav-cor} \rangle$

(d) $\langle \Delta H_{bind} \rangle$

Figure 5: hydroxyaniline



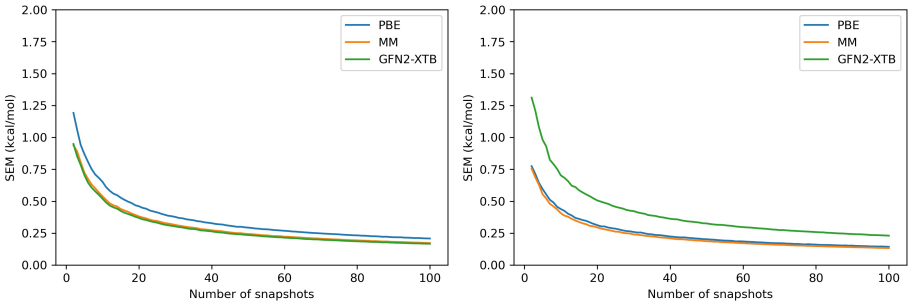
(a) $\langle \Delta E \rangle$

(b) $\langle \Delta G_{solvation} \rangle$

(c) $\langle \Delta G_{solvation-cav-cor} \rangle$

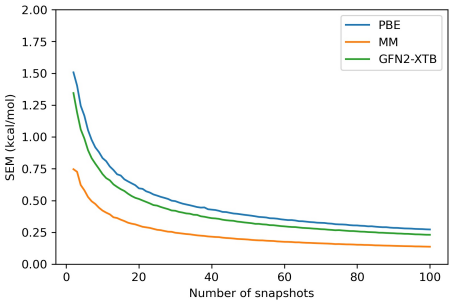
(d) $\langle \Delta H_{bind} \rangle$

Figure 6: toluene

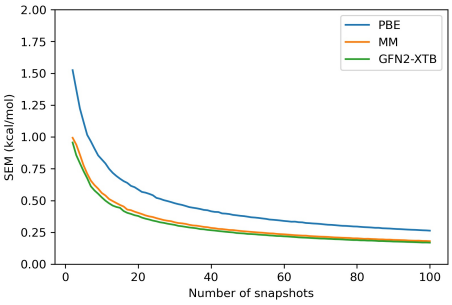


(a) $\langle \Delta E \rangle$

(b) $\langle \Delta G_{solvation} \rangle$



(c) $\langle \Delta G_{solvation-cav-cor} \rangle$



(d) $\langle \Delta H_{bind} \rangle$

Figure 7: chlorophenol

2.1.2 Entropic components

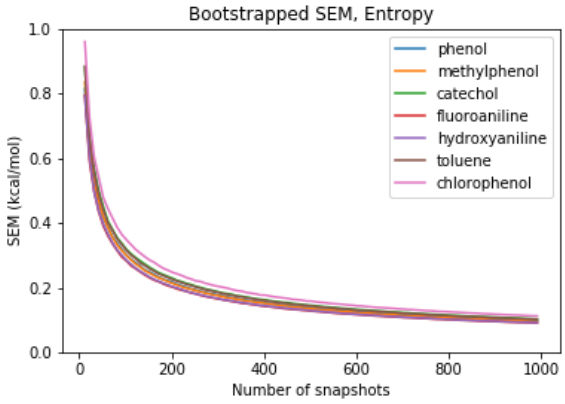


Figure 8: Bootstrapped (1000 re-samples) SEM of entropy correction term, $T \times \langle \Delta S \rangle$, up to 100 snapshots for each ligand.

2.2 Mean enthalpy change and absolute deviations

2.2.1 Equally Spaced Snapshots

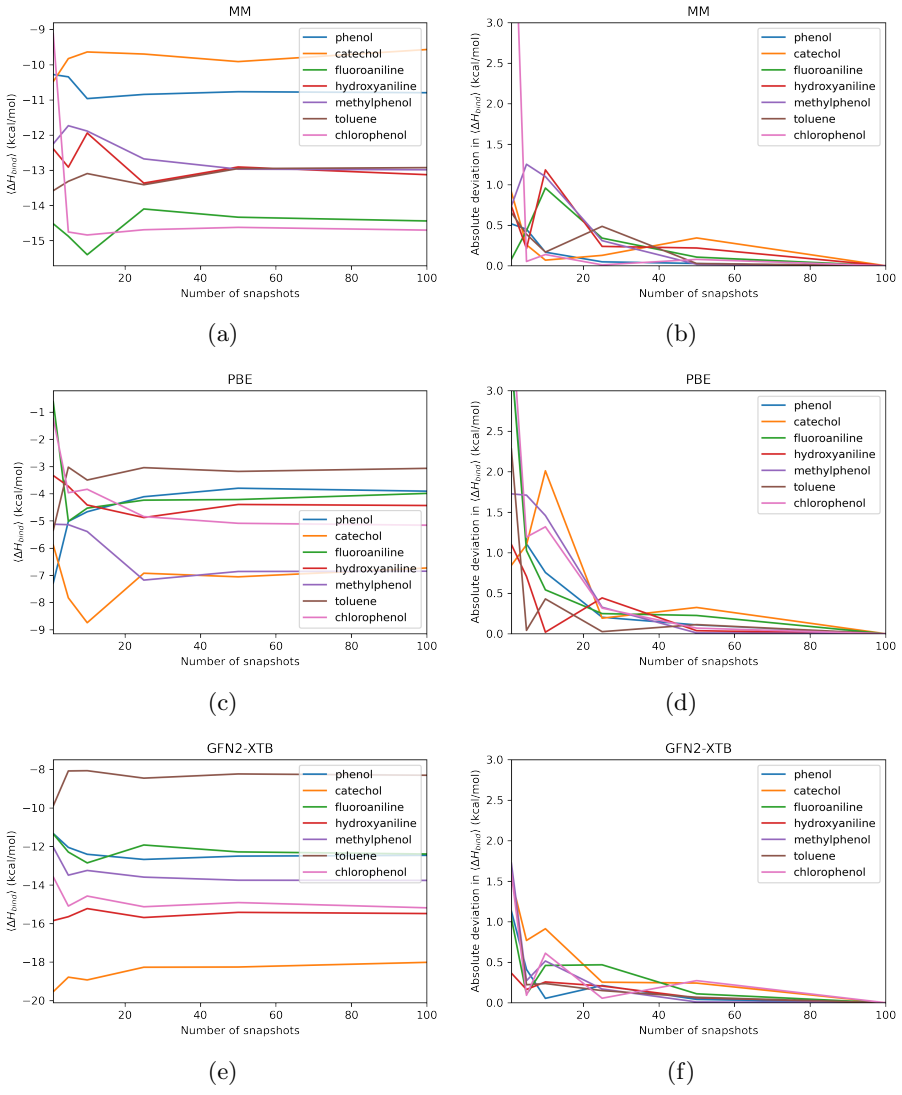


Figure 9: Left: Mean change in total enthalpy upon binding,$\langle \Delta H_{bind} \rangle$, of each ligand at different numbers of equally spaced snapshots. Right: Absolute deviation of $\langle \Delta H_{bind} \rangle$ at different numbers of equally spaced snapshots from the 'converged' mean over 100 snapshots. Methods: MM (a,b), DFT(PBE) (c,d), GFN2-XTB (e,f).

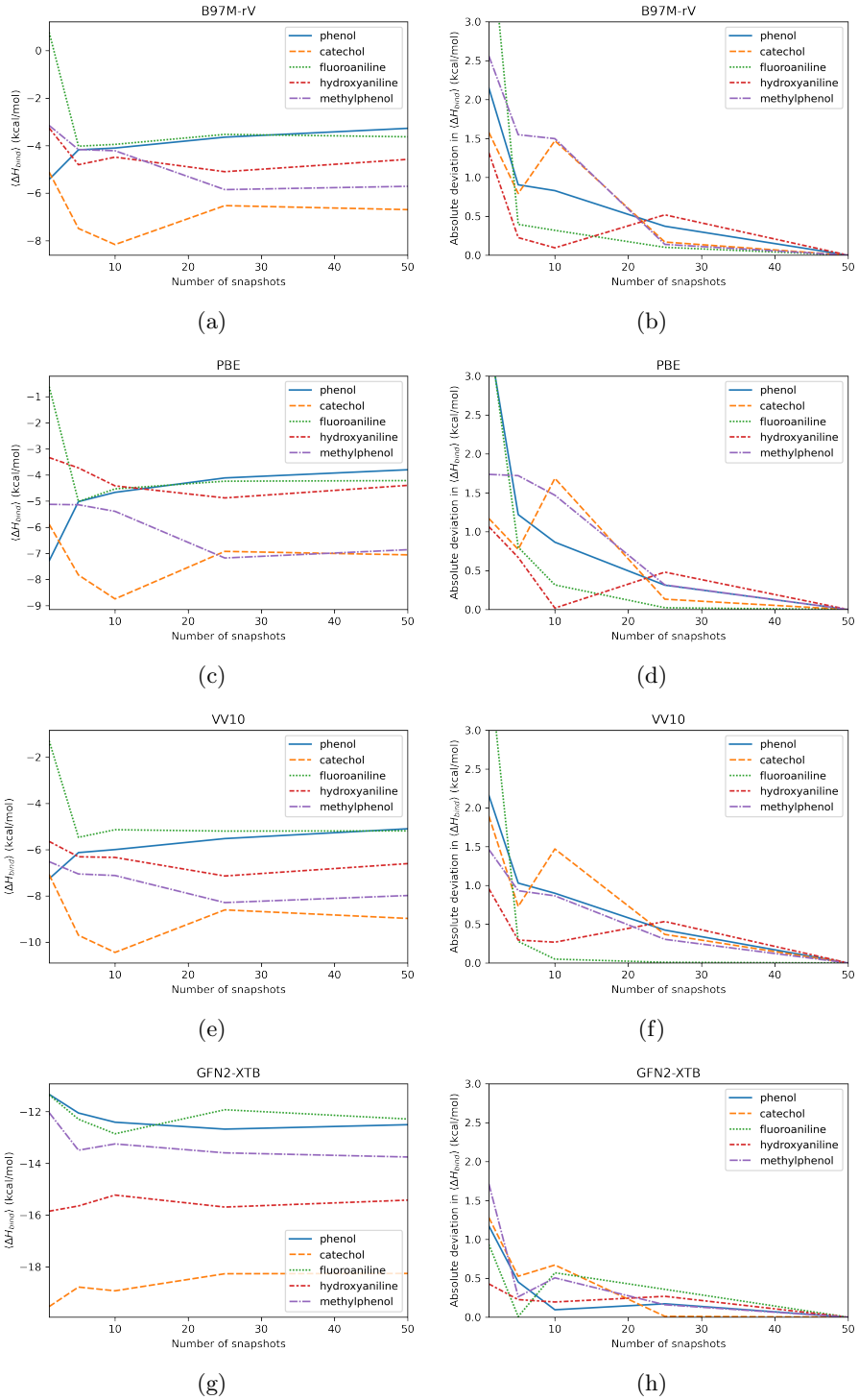


Figure 10: Left: Mean change in total enthalpy upon binding,$\langle \Delta H_{bind} \rangle$, of each ligand at different numbers of equally spaced snapshots. Right: Absolute deviation of $\langle \Delta H_{bind} \rangle$ at different numbers of equally spaced snapshots from the 'converged' mean over 50 snapshots. Methods: B97M-rV (a,b),PBE (c,d), VV10 (e,f), GFN2-XTB (g,h).

2.2.2 Randomly selected Snapshots

Figures 11-17: Absolute deviation of $\langle \Delta H_{bind} \rangle$ at different numbers of randomly selected snapshots from the 'converged' mean over 100 snapshots. Two different sets of random snapshots shown in blue and orange.

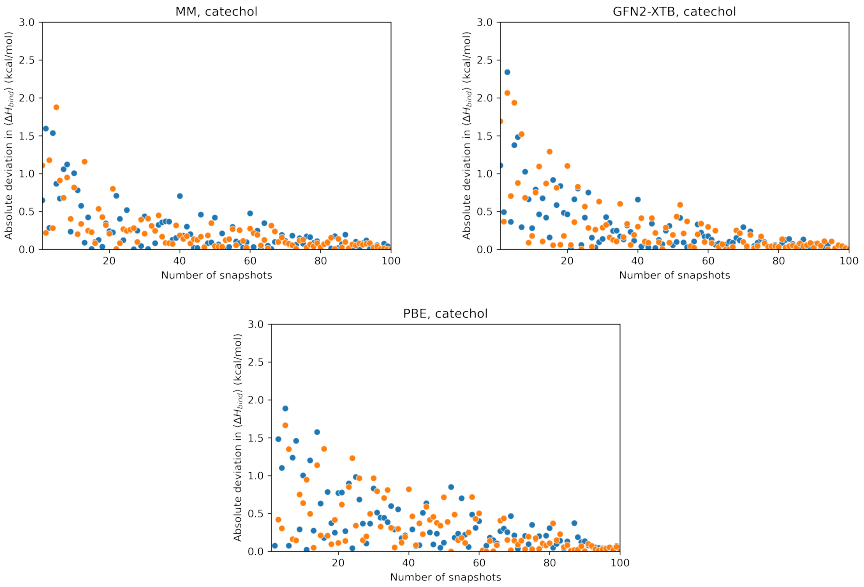


Figure 11: catechol

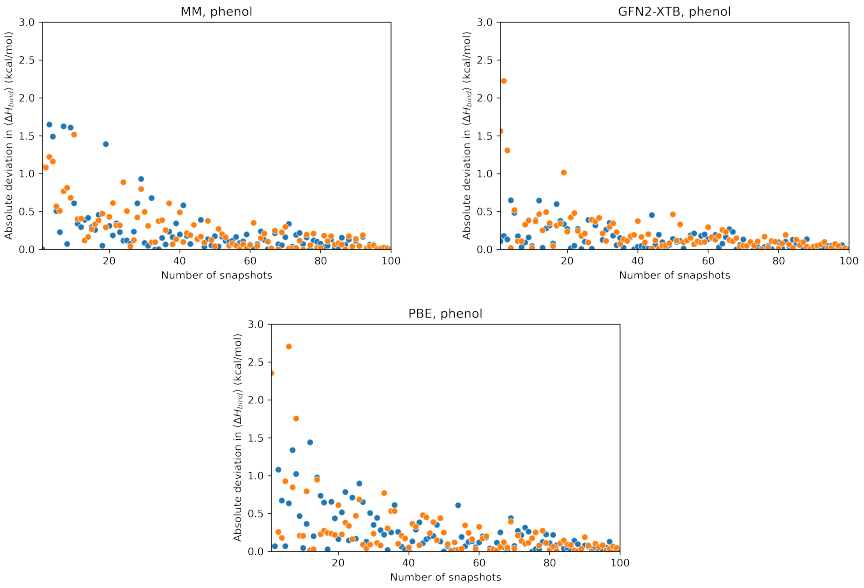


Figure 12: phenol

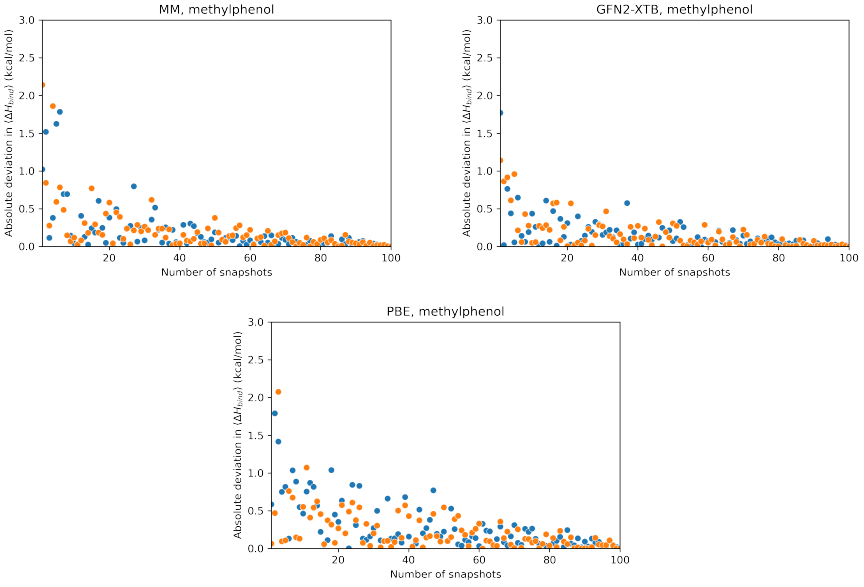


Figure 13: methylphenol

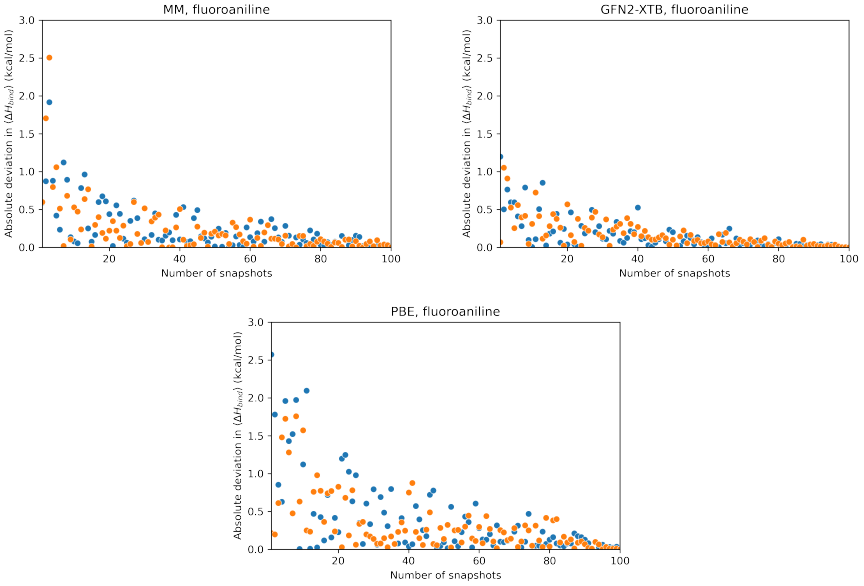


Figure 14: fluoroaniline

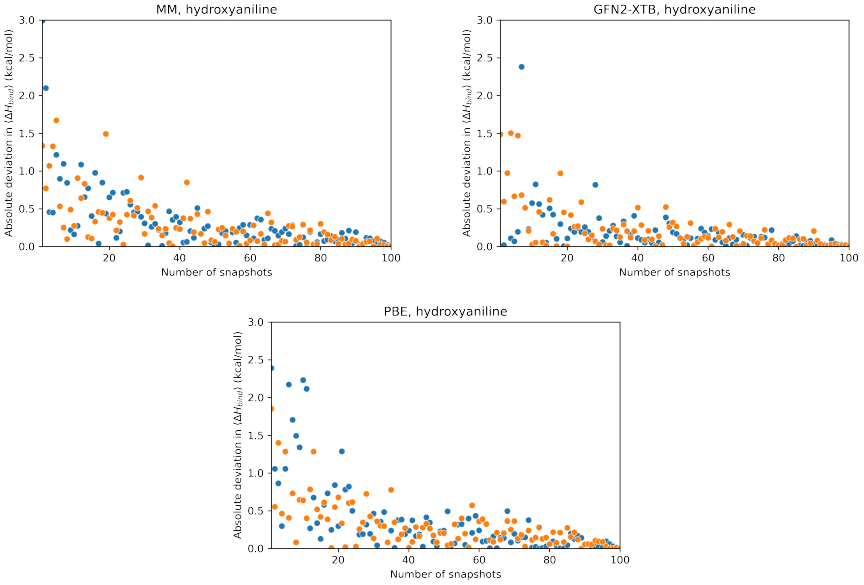


Figure 15: hydroxyaniline

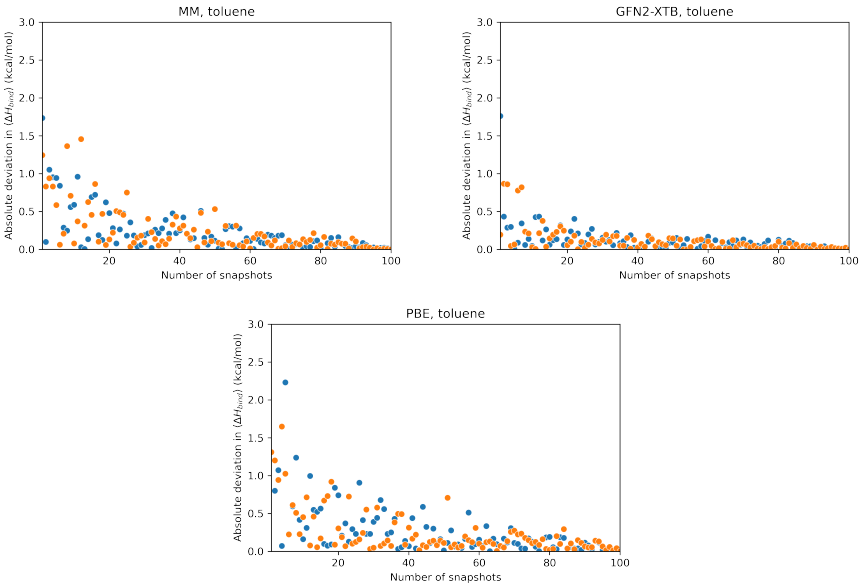


Figure 16: toluene

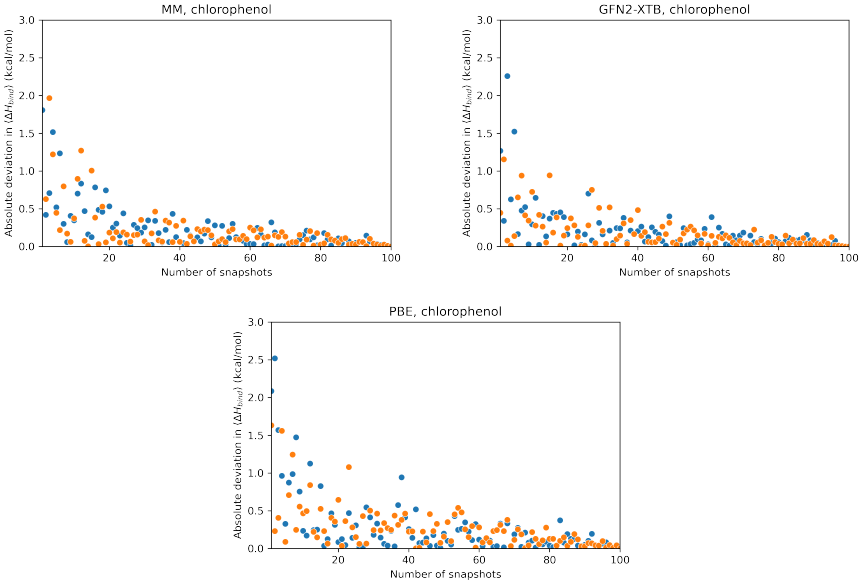


Figure 17: chlorophenol

2.3 RMSDtr with Entropy Sampling

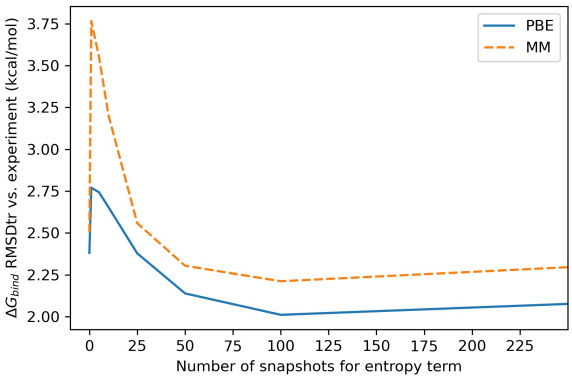


Figure 18: Root mean square error deviation from experiment after removal of mean signed error (kcal/mol) of calculated binding free energies for ligand set B, at different levels of entropy sampling. Enthalpy sampled over 100 snapshots. RMSDtr calculated with upper limit for non-binder.

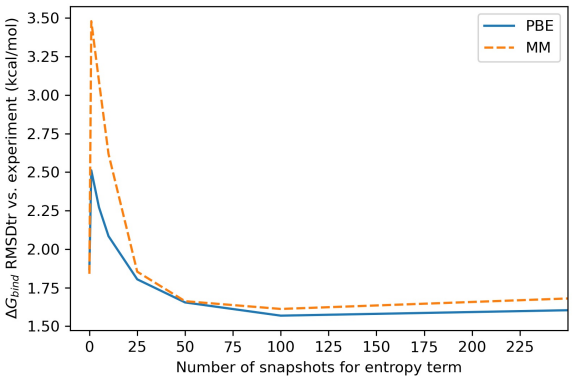


Figure 19: Root mean square error deviation from experiment after removal of mean signed error (kcal/mol) of calculated binding free energies for ligand set B, at different levels of entropy sampling. Enthalpy sampled over 100 snapshots. RMSDtr calculated with lower limit for non-binder.

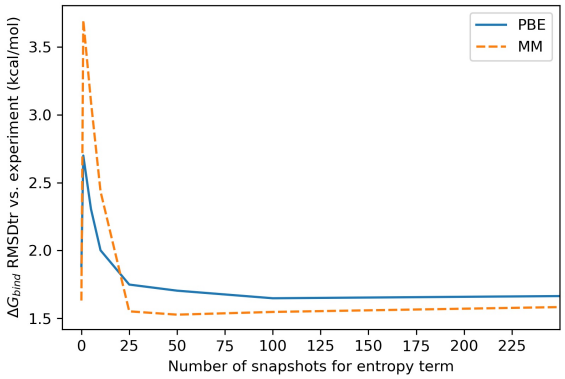


Figure 20: Root mean square error deviation from experiment after removal of mean signed error (kcal/mol) of calculated binding free energies for ligand set B, at different levels of entropy sampling. Enthalpy sampled over 100 snapshots. RMSDtr calculated with binders only.

2.4 RMSDtr convergence

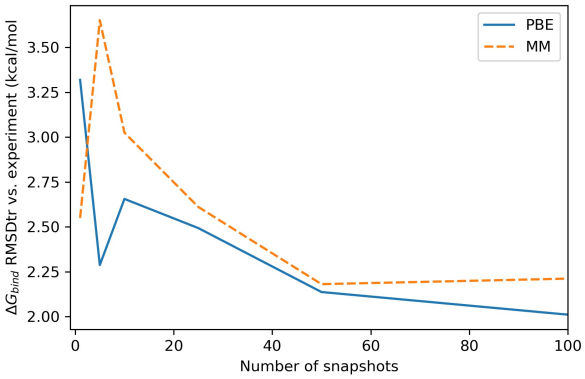


Figure 21: Root mean square error deviation from experiment after removal of mean signed error (kcal/mol) of calculated binding free energies for ligand set B, at different levels of entropy and energy sampling. Enthalpy and energy sampled over same snapshots. RMSDtr calculated with upper limit for non-binder.

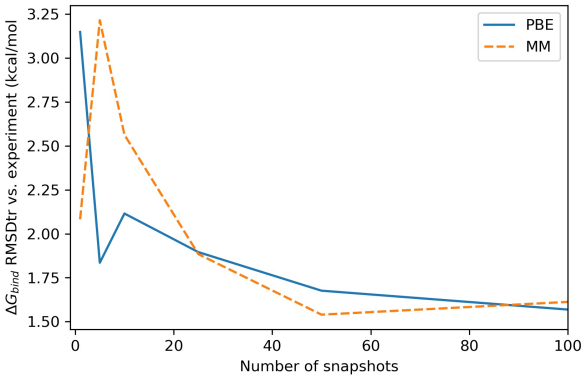


Figure 22: Root mean square error deviation from experiment after removal of mean signed error (kcal/mol) of calculated binding free energies for ligand set B, at different levels of entropy and energy sampling. Enthalpy and energy sampled over same snapshots. RMSDtr calculated with lower limit for non-binder.

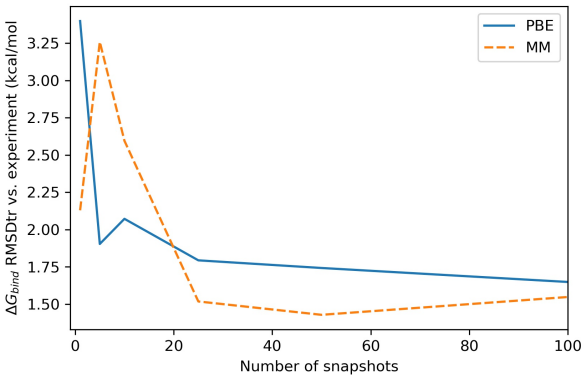
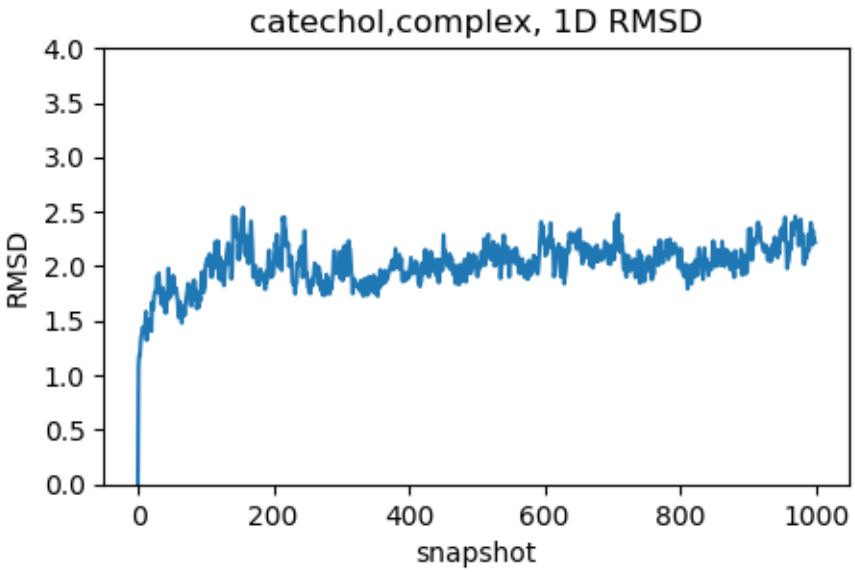


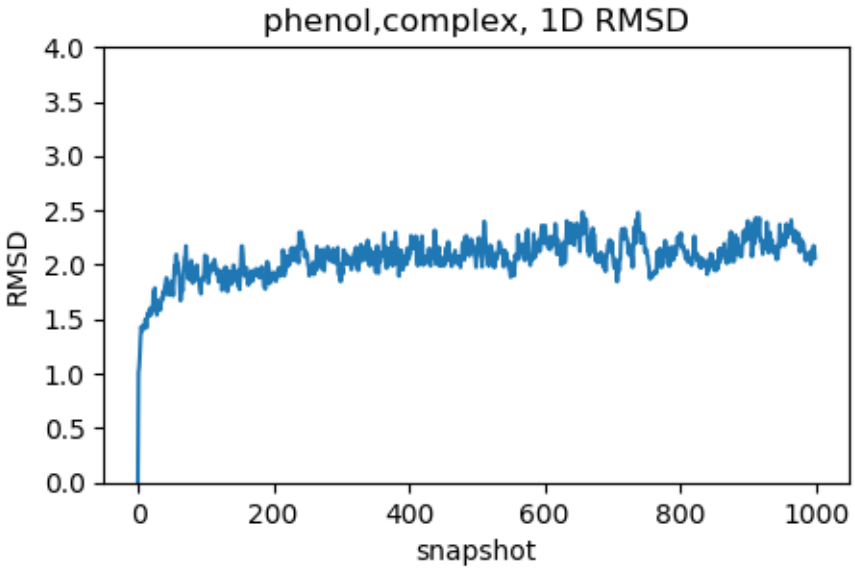
Figure 23: Root mean square error deviation from experiment after removal of mean signed error (kcal/mol) of calculated binding free energies for ligand set B, at different levels of entropy and energy sampling. Enthalpy and energy sampled over same snapshots. RMSDtr calculated with binders only.

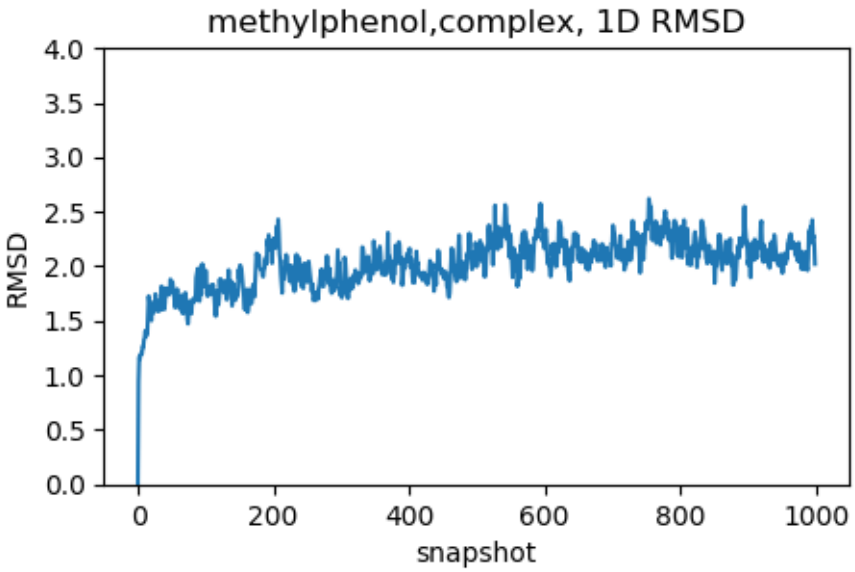
2.5 Analysis of Trajectories

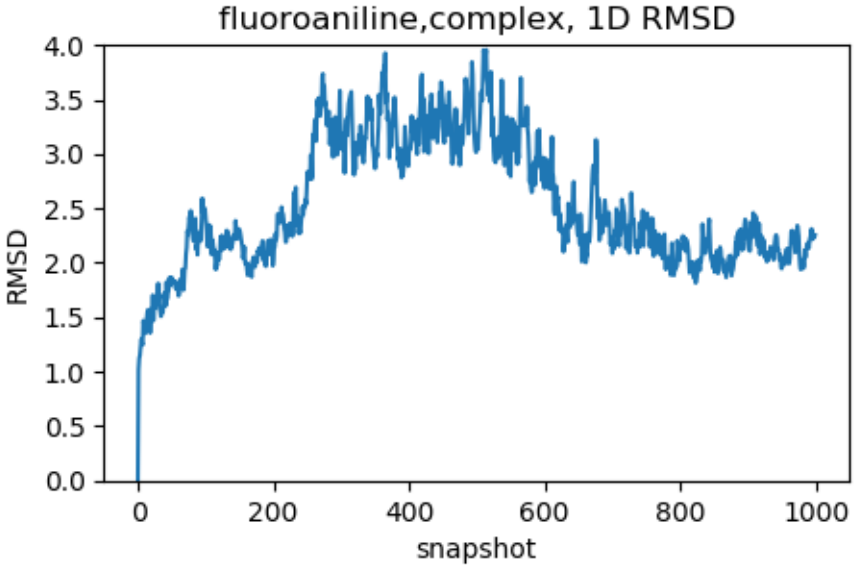
2.5.1 1D RSMD

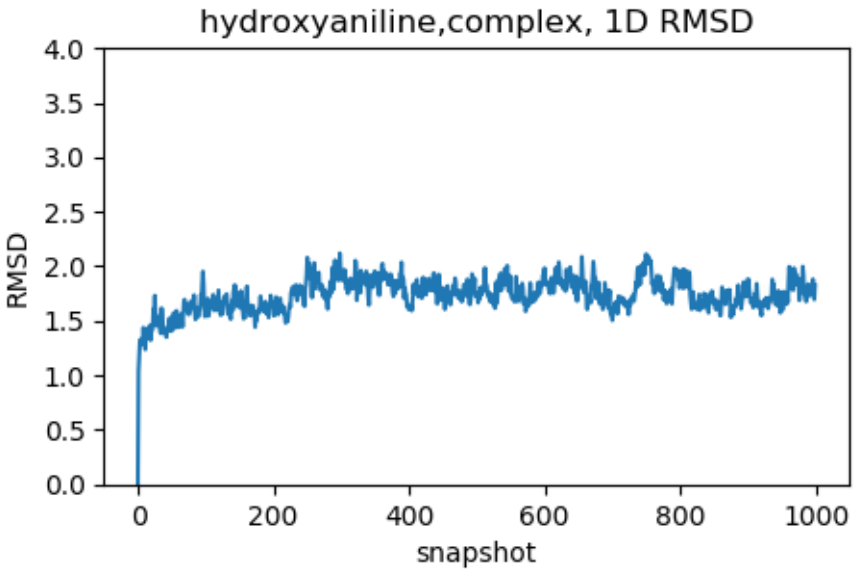
One dimensional RMSD (Ångstrom) plots with respect to first snapshot of MD trajectory for each ligand in complex with (L99A/M102Q).

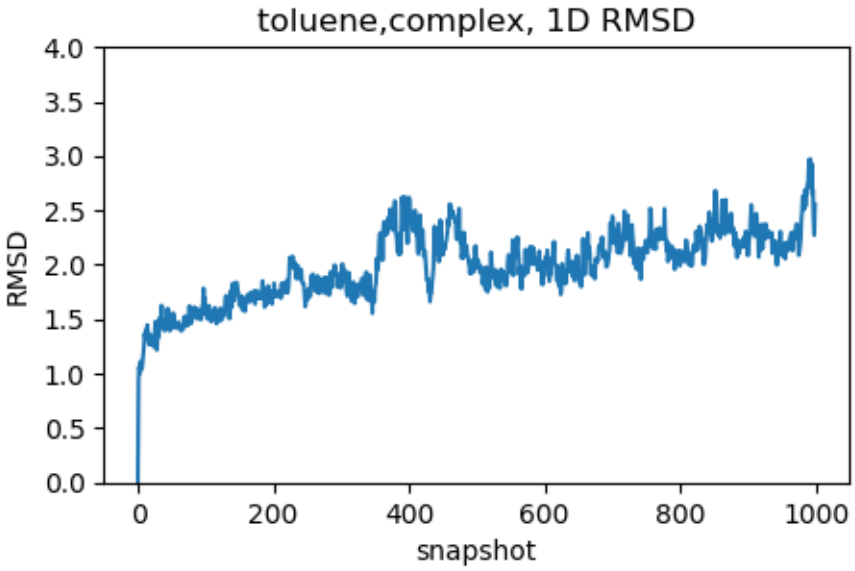

catechol,complex, 1D RMSD
RMSD
snapshot
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
200
400
600
800
1000

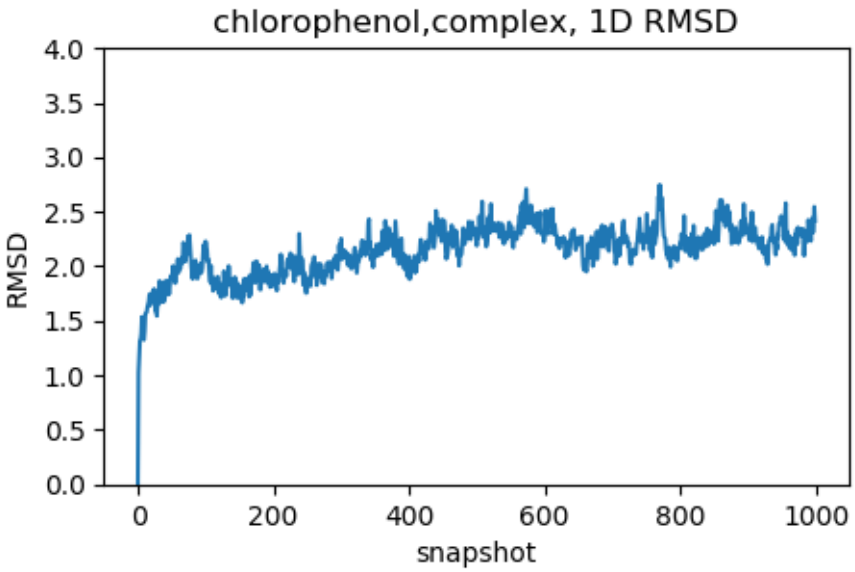

phenol,complex, 1D RMSD
RMSD
snapshot
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
200
400
600
800
1000


methylphenol,complex, 1D RMSD
RMSD
snapshot
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
200
400
600
800
1000


fluoroaniline,complex, 1D RMSD
RMSD
snapshot
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
200
400
600
800
1000

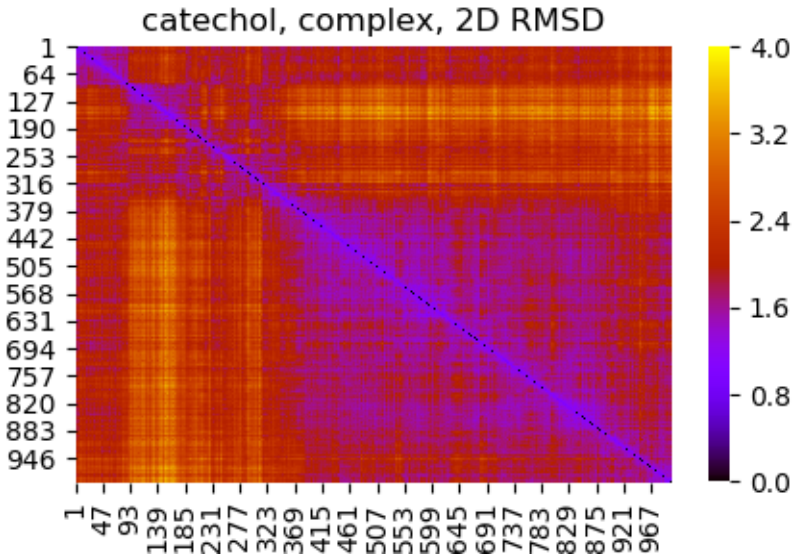

hydroxyaniline,complex, 1D RMSD
RMSD
snapshot
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
200
400
600
800
1000

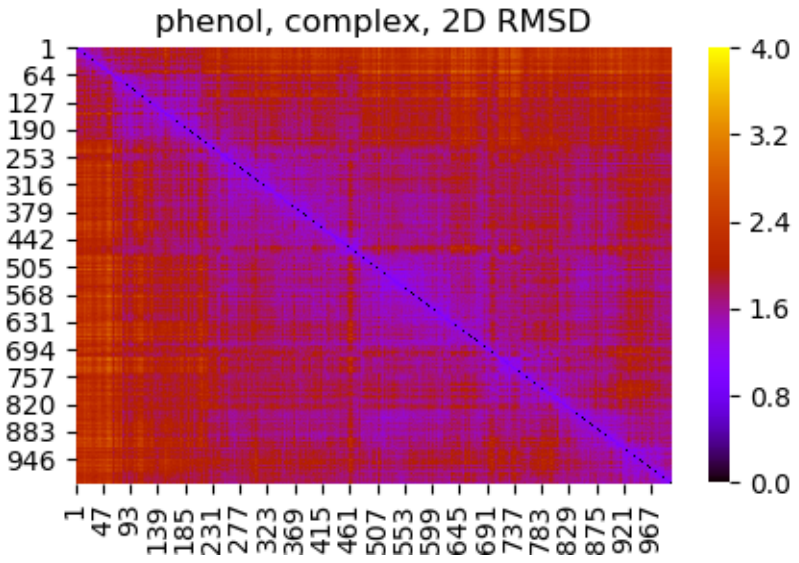

toluene,complex, 1D RMSD
RMSD
snapshot
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
200
400
600
800
1000

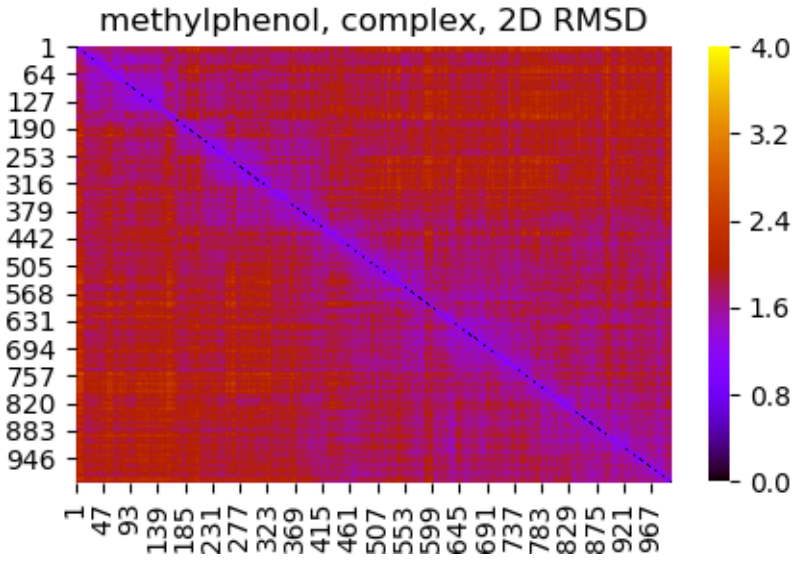


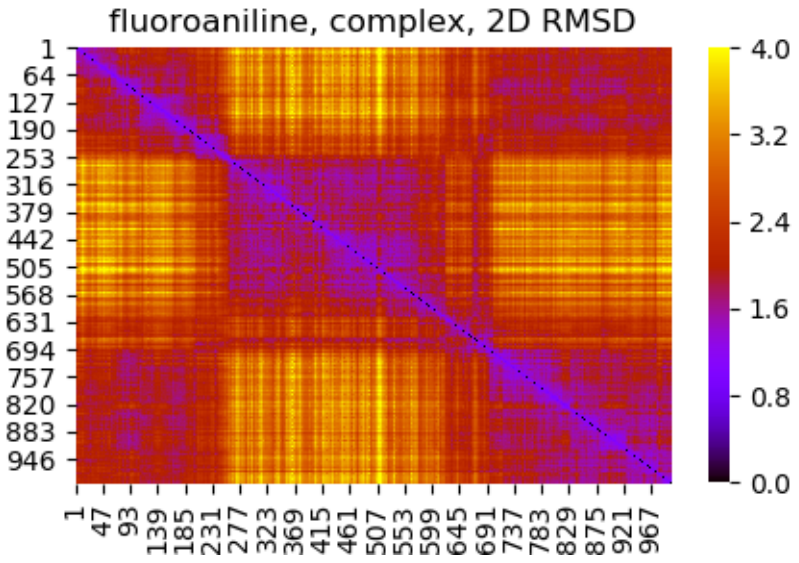
2.5.2 2D RSMD

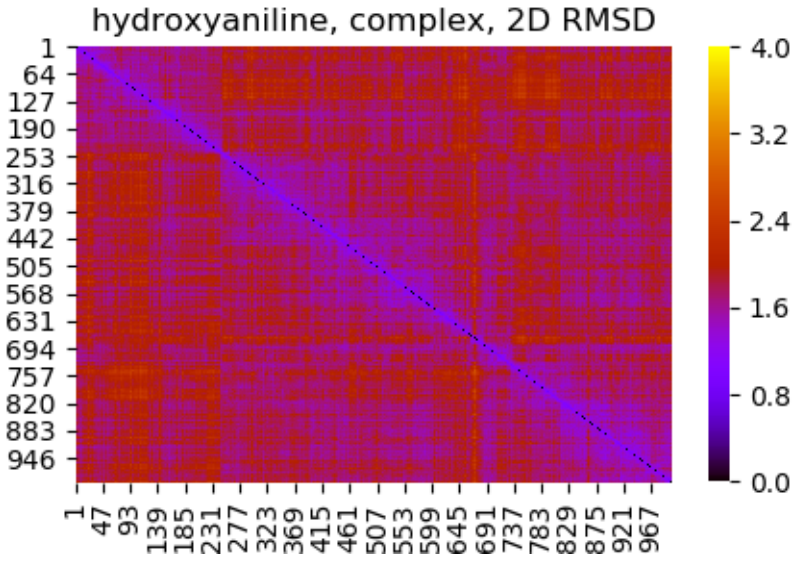
Heatmaps of 2D RMSD (Ångstrom) of MD trajectories for each ligand in complex with T4-lysozyme (L99A/M102Q).

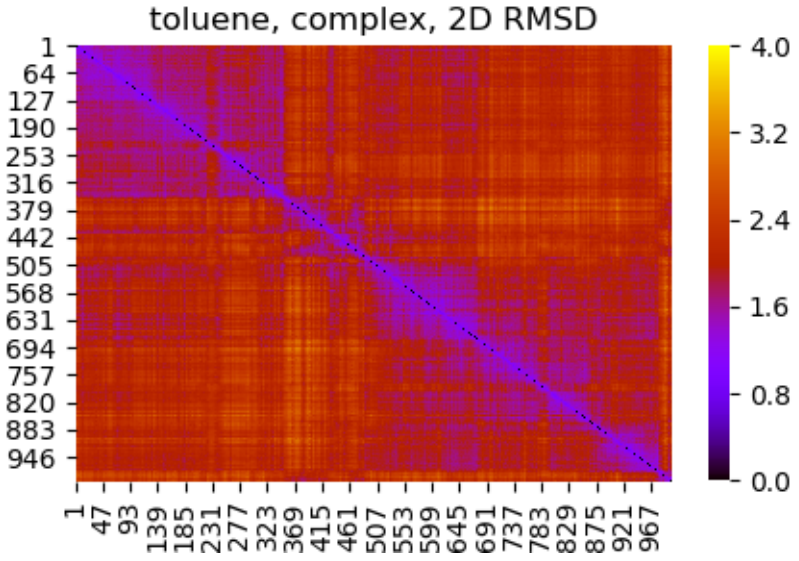


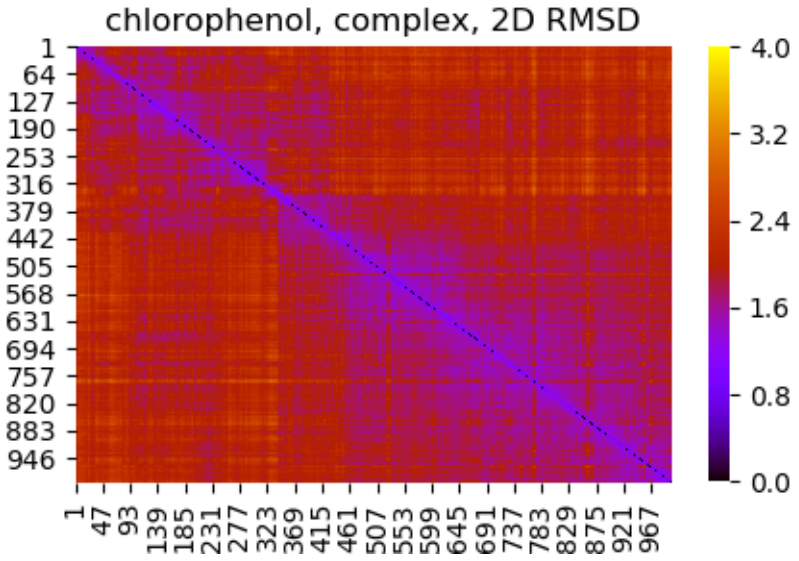

phenol, complex, 2D RMSD
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1
47
93
139
185
231
277
323
369
415
461
507
553
599
645
691
737
783
829
875
921
967
4.0
3.2
2.4
1.6
0.8
0.0


methylphenol, complex, 2D RMSD
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1
47
93
139
185
231
277
323
369
415
461
507
553
599
645
691
737
783
829
875
921
967
4.0
3.2
2.4
1.6
0.8
0.0


fluoroaniline, complex, 2D RMSD
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1
47
93
139
185
231
277
323
369
415
461
507
553
599
645
691
737
783
829
875
921
967
4.0
3.2
2.4
1.6
0.8
0.0


hydroxyaniline, complex, 2D RMSD
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1
47
93
139
185
231
277
323
369
415
461
507
553
599
645
691
737
783
829
875
921
967
4.0
3.2
2.4
1.6
0.8
0.0


toluene, complex, 2D RMSD
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1
47
93
139
185
231
277
323
369
415
461
507
553
599
645
691
737
783
829
875
921
967
4.0
3.2
2.4
1.6
0.8
0.0


chlorophenol, complex, 2D RMSD
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1
47
93
139
185
231
277
323
369
415
461
507
553
599
645
691
737
783
829
875
921
967
4.0
3.2
2.4
1.6
0.8
0.0

2.5.3 RMSD filter

RMSE with and without RMSD filter

Reference Ligand	phenol	catechol	fluoroaniline	methylphenol	toluene	chlorophenol
PBE	2.79	3.09	3.19	2.57	3.36	2.30
PBE_3.0	2.59	2.86	3.02	2.49	3.28	2.20
MM	3.53	4.22	2.53	2.56	2.54	2.86
MM_3.0	3.39	4.41	2.54	2.55	2.53	2.85

Table 1: Root mean square error (RMSE) of entropy-corrected binding free energies relative to each reference ligand against experiment with and without a 3 Ångstrom RMSD filter for PBE and MM at 100 snapshots (Ligand set B).

2.6 Energy Components

Each table contains the main energy and entropy components for each ligand. The gas-phase energy E contains the empirical dispersion correction Emp_Disp. G_bind with and without entropy term given. Entropy term is $T\Delta S$.

2.6.1 Ligand set A

Phenol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-28.335	3.143	24.535	-3.800	-22.304	-14.219	10.419
PBE_bj	-26.541	3.143	24.535	-2.006	-20.511	-14.219	12.213
PBE_bjm	-25.774	3.143	24.535	-1.240	-19.744	-14.219	12.979
PBE_zero	-27.892	3.143	24.535	-3.358	-21.862	-14.219	10.862
PBE_old	-26.397	3.143	24.535	-1.862	-20.366	-14.219	12.357
PBE_zerom	-24.616	3.143	24.535	-0.081	-18.585	-14.219	14.138
PBE_elstner	-28.335	3.143	24.535	-3.800	-22.304	-14.219	10.419
PBE_bj_abc	-24.364	3.143	24.535	0.170	-18.334	-14.219	14.389
PBE_bjm_abc	-23.598	3.143	24.535	0.937	-17.567	-14.219	15.156
B97M-V	-29.665	3.096	26.395	-3.270	0.000	-14.219	10.950
VV10	-30.825	2.812	25.730	-5.095	0.000	-14.219	9.124
MM	-25.834	14.491	15.066	-10.769	0.000	-14.219	3.450

Catechol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-36.851	7.040	29.793	-7.057	-23.413	-14.615	7.558
PBE_bj	-35.236	7.040	29.793	-5.443	-21.799	-14.615	9.173
PBE_bjm	-34.433	7.040	29.793	-4.639	-20.995	-14.615	9.976
PBE_zero	-36.769	7.040	29.793	-6.976	-23.332	-14.615	7.640
PBE_old	-34.963	7.040	29.793	-5.170	-21.526	-14.615	9.446
PBE_zerom	-33.200	7.040	29.793	-3.407	-19.763	-14.615	11.208
PBE_elstner	-36.851	7.040	29.793	-7.057	-23.413	-14.615	7.558
PBE_bj_abc	-33.028	7.040	29.793	-3.235	-19.591	-14.615	11.381
PBE_bjm_abc	-32.224	7.040	29.793	-2.431	-18.787	-14.615	12.184
B97M-V	-38.599	7.179	31.907	-6.692	0.000	-14.615	7.923
VV10	-40.127	6.737	31.156	-8.970	0.000	-14.615	5.645
MM	-36.297	25.792	26.384	-9.913	0.000	-14.615	4.702

Methylphenol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-30.854	2.304	23.993	-6.860	-24.120	-15.772	8.911
PBE_bj	-29.026	2.304	23.993	-5.032	-22.292	-15.772	10.739
PBE_bjm	-28.067	2.304	23.993	-4.074	-21.334	-15.772	11.698
PBE_zero	-30.732	2.304	23.993	-6.738	-23.998	-15.772	9.034
PBE_old	-28.649	2.304	23.993	-4.656	-21.915	-15.772	11.116
PBE_zerom	-26.850	2.304	23.993	-2.857	-20.117	-15.772	12.915
PBE_elstner	-30.854	2.304	23.993	-6.860	-24.120	-15.772	8.911
PBE_bj_abc	-26.645	2.304	23.993	-2.652	-19.912	-15.772	13.120
PBE_bjm_abc	-25.687	2.304	23.993	-1.693	-18.953	-15.772	14.079
B97M-V	-32.750	2.793	27.039	-5.711	0.000	-15.772	10.061
VV10	-33.642	1.792	25.661	-7.981	0.000	-15.772	7.791
MM	-28.466	14.737	15.498	-12.968	0.000	-15.772	2.803

Fluoroaniline	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-27.042	2.075	22.826	-4.216	-22.363	-15.374	11.158
PBE_bj	-25.296	2.075	22.826	-2.469	-20.617	-15.374	12.905
PBE_bjm	-24.452	2.075	22.826	-1.625	-19.773	-15.374	13.749
PBE_zero	-26.895	2.075	22.826	-4.069	-22.216	-15.374	11.305
PBE_old	-25.354	2.075	22.826	-2.528	-20.675	-15.374	12.846
PBE_zerom	-23.327	2.075	22.826	-0.500	-18.648	-15.374	14.874
PBE_elstner	-27.042	2.075	22.826	-4.216	-22.363	-15.374	11.158
PBE_bj_abc	-23.191	2.075	22.826	-0.365	-18.512	-15.374	15.009
PBE_bjm_abc	-22.347	2.075	22.826	0.480	-17.668	-15.374	15.854
B97M-V	-28.669	2.135	25.043	-3.626	0.000	-15.374	11.748
VV10	-30.085	2.235	24.899	-5.186	0.000	-15.374	10.188
MM	-26.011	10.980	11.677	-14.334	0.000	-15.374	1.040

Hydroxyaniline	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-34.885	7.432	30.488	-4.398	-25.196	-15.232	10.835
PBE_bj	-32.997	7.432	30.488	-2.509	-23.308	-15.232	12.723
PBE_bjm	-32.088	7.432	30.488	-1.600	-22.399	-15.232	13.632
PBE_zero	-34.552	7.432	30.488	-4.064	-24.862	-15.232	11.168
PBE_old	-33.151	7.432	30.488	-2.663	-23.461	-15.232	12.570
PBE_zerom	-30.770	7.432	30.488	-0.282	-21.080	-15.232	14.951
PBE_elstner	-34.885	7.432	30.488	-4.398	-25.196	-15.232	10.835
PBE_bj_abc	-30.605	7.432	30.488	-0.118	-20.916	-15.232	15.115
PBE_bjm_abc	-29.696	7.432	30.488	0.791	-20.007	-15.232	16.024
B97M-V	-37.042	7.455	32.466	-4.576	0.000	-15.232	10.656
VV10	-38.073	6.881	31.471	-6.601	0.000	-15.232	8.631
MM	-33.515	19.984	20.608	-12.907	0.000	-15.232	2.326

2.6.2 Ligand Set B

Phenol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-28.285	3.251	24.375	-3.909	-22.276	-14.213	10.304
PBE_bj	-26.504	3.251	24.375	-2.128	-20.495	-14.213	12.085
PBE_bjm	-25.728	3.251	24.375	-1.352	-19.719	-14.213	12.861
PBE_zero	-27.867	3.251	24.375	-3.491	-21.858	-14.213	10.722
PBE_old	-26.359	3.251	24.375	-1.984	-20.351	-14.213	12.229
PBE_zerom	-24.570	3.251	24.375	-0.195	-18.561	-14.213	14.018
PBE_elstner	-28.285	3.251	24.375	-3.909	-22.276	-14.213	10.304
PBE_bj_abc	-24.382	3.251	24.375	-0.006	-18.373	-14.213	14.207
PBE_bjm_abc	-23.606	3.251	24.375	0.770	-17.597	-14.213	14.982
MM	-25.841	14.471	15.045	-10.797	0.000	-14.213	3.416
GFN2-XTB	-16.241	7.257	7.257	-12.463	0.000	-14.213	1.750
GFN-FF	-21.721	11.993	11.993	-18.152	0.000	-14.213	-3.940

Catechol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-36.454	7.006	29.721	-6.733	-23.467	-14.782	8.049
PBE_bj	-34.836	7.006	29.721	-5.115	-21.849	-14.782	9.667
PBE_bjm	-34.040	7.006	29.721	-4.318	-21.053	-14.782	10.463
PBE_zero	-36.350	7.006	29.721	-6.628	-23.363	-14.782	8.153
PBE_old	-34.655	7.006	29.721	-4.934	-21.668	-14.782	9.848
PBE_zerom	-32.799	7.006	29.721	-3.078	-19.812	-14.782	11.704
PBE_elstner	-36.454	7.006	29.721	-6.733	-23.467	-14.782	8.049
PBE_bj_abc	-32.631	7.006	29.721	-2.909	-19.643	-14.782	11.872
PBE_bjm_abc	-31.834	7.006	29.721	-2.113	-18.847	-14.782	12.669
MM	-35.770	25.612	26.201	-9.570	0.000	-14.782	5.212
GFN2-XTB	-22.762	9.611	9.611	-18.013	0.000	-14.782	-3.232
GFN-FF	-26.716	16.400	16.400	-22.162	0.000	-14.782	-7.380

Methylphenol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-30.811	2.255	23.961	-6.850	-24.130	-15.459	8.609
PBE_bj	-28.986	2.255	23.961	-5.025	-22.305	-15.459	10.435
PBE_bjm	-28.031	2.255	23.961	-4.069	-21.350	-15.459	11.390
PBE_zero	-30.697	2.255	23.961	-6.736	-24.016	-15.459	8.723
PBE_old	-28.633	2.255	23.961	-4.672	-21.952	-15.459	10.788
PBE_zerom	-26.813	2.255	23.961	-2.852	-20.132	-15.459	12.608
PBE_elstner	-30.811	2.255	23.961	-6.850	-24.130	-15.459	8.609
PBE_bj_abc	-26.612	2.255	23.961	-2.650	-19.931	-15.459	12.809
PBE_bjm_abc	-25.656	2.255	23.961	-1.695	-18.975	-15.459	13.764
MM	-28.478	14.738	15.491	-12.986	0.000	-15.459	2.473
GFN2-XTB	-17.800	7.362	7.362	-13.763	0.000	-15.459	1.696
GFN-FF	-23.628	12.281	12.281	-19.670	0.000	-15.459	-4.211

Fluoroaniline	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-27.287	2.401	23.296	-3.990	-22.554	-15.142	11.151
PBE_bj	-25.497	2.401	23.296	-2.200	-20.764	-15.142	12.941
PBE_bjm	-24.656	2.401	23.296	-1.359	-19.923	-15.142	13.782
PBE_zero	-27.103	2.401	23.296	-3.806	-22.370	-15.142	11.335
PBE_old	-25.620	2.401	23.296	-2.323	-20.887	-15.142	12.818
PBE_zerom	-23.518	2.401	23.296	-0.221	-18.785	-15.142	14.920
PBE_elstner	-27.287	2.401	23.296	-3.990	-22.554	-15.142	11.151
PBE_bj_abc	-23.382	2.401	23.296	-0.086	-18.650	-15.142	15.056
PBE_bjm_abc	-22.541	2.401	23.296	0.755	-17.809	-15.142	15.897
MM	-26.212	11.083	11.772	-14.440	0.000	-15.142	0.701
GFN2-XTB	-15.508	7.807	7.807	-12.395	0.000	-15.142	2.747
GFN-FF	-21.940	17.776	17.776	-17.535	0.000	-15.142	-2.393

Hydroxyaniline	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-35.076	7.526	30.641	-4.436	-25.267	-15.576	11.140
PBE_bj	-33.171	7.526	30.641	-2.530	-23.362	-15.576	13.046
PBE_bjm	-32.265	7.526	30.641	-1.625	-22.457	-15.576	13.951
PBE_zero	-34.720	7.526	30.641	-4.079	-24.911	-15.576	11.497
PBE_old	-33.359	7.526	30.641	-2.718	-23.550	-15.576	12.858
PBE_zerom	-30.942	7.526	30.641	-0.301	-21.133	-15.576	15.275
PBE_elstner	-35.076	7.526	30.641	-4.436	-25.267	-15.576	11.140
PBE_bj_abc	-30.771	7.526	30.641	-0.131	-20.963	-15.576	15.445
PBE_bjm_abc	-29.866	7.526	30.641	0.774	-20.057	-15.576	16.350
MM	-33.652	19.911	20.527	-13.126	0.000	-15.576	2.450
GFN2-XTB	-20.122	11.469	11.469	-15.483	0.000	-15.576	0.094
GFN-FF	-27.882	18.434	18.434	-22.752	0.000	-15.576	-7.176

Toluene	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-21.028	-2.952	17.958	-3.070	-23.794	-14.389	11.319
PBE_bj	-18.596	-2.952	17.958	-0.638	-21.362	-14.389	13.751
PBE_bjm	-17.704	-2.952	17.958	0.254	-20.470	-14.389	14.643
PBE_zero	-20.043	-2.952	17.958	-2.084	-22.808	-14.389	12.305
PBE_old	-18.604	-2.952	17.958	-0.646	-21.370	-14.389	13.743
PBE_zerom	-16.540	-2.952	17.958	1.418	-19.305	-14.389	15.807
PBE_elstner	-21.028	-2.952	17.958	-3.070	-23.794	-14.389	11.319
PBE_bj_abc	-16.431	-2.952	17.958	1.527	-19.197	-14.389	15.916
PBE_bjm_abc	-15.539	-2.952	17.958	2.419	-18.305	-14.389	16.808
MM	-19.244	5.566	6.319	-12.925	0.000	-14.389	1.464
GFN2-XTB	-10.713	3.070	3.070	-8.309	0.000	-14.389	6.080
GFN-FF	-18.388	8.154	8.154	-15.696	0.000	-14.389	-1.307

Chlorophenol	E	G_solv	G_solv_corrected	G_bind	Emp_Disp	Entropy	G_bind_with_Entropy
PBE+ONETEP_DISP	-31.278	3.044	26.119	-5.159	-24.061	-14.131	8.972
PBE_bj	-30.218	3.044	26.119	-4.099	-23.000	-14.131	10.032
PBE_bjm	-29.643	3.044	26.119	-3.524	-22.425	-14.131	10.607
PBE_zero	-31.696	3.044	26.119	-5.576	-24.478	-14.131	8.554
PBE_old	-29.630	3.044	26.119	-3.511	-22.412	-14.131	10.620
PBE_zerom	-28.172	3.044	26.119	-2.052	-20.954	-14.131	12.078
PBE_elstner	-31.278	3.044	26.119	-5.159	-24.061	-14.131	8.972
PBE_bj_abc	-27.889	3.044	26.119	-1.770	-20.671	-14.131	12.361
PBE_bjm_abc	-27.314	3.044	26.119	-1.195	-20.096	-14.131	12.936
MM	-29.617	14.223	14.917	-14.701	0.000	-14.131	-0.570
GFN2-XTB	-19.408	6.991	6.991	-15.185	0.000	-14.131	-1.055
GFN-FF	-25.781	12.126	12.126	-21.559	0.000	-14.131	-7.428

References

[1]J. C. Prentice, J. Aarons, J. C. Womack, A. E. Allen, L. Andrinopoulos, L. Anton, R. A. Bell, A. Bhandari, G. A. Bramley, R. J. Charlton, R. J. Clements, D. J. Cole, G. Constantinescu, F. Corsetti, S. M. Dubois, K. K. Duff, J. M. Escartín, A. Greco, Q. Hill, L. P. Lee, E. Linscott, D. D. O'Regan, M. J. Phipps, L. E. Ratcliff, Á. R. Serrano, E. W. Tait, G. Teobaldi, V. Vitale, N. Yeung, T. J. Zuehlsdorff, J. Dziedzic, P. D. Haynes, N. D. Hine, A. A. Mostofi, M. C. Payne, and C. K. Skylaris, "The ONETEP linear-scaling density functional theory program", The Journal of Chemical Physics **152**, 174111 (2020).

[2]J. Dziedzic, H. H. Helal, C. K. Skylaris, A. A. Mostofi, and M. C. Payne, "Minimal parameter implicit solvent model for ab initio electronic-structure calculations", EPL **95**, 43001 (2011).

[3]S. Genheden, J. Kongsted, P. Söderhjelm, and U. Ryde, "Nonpolar solvation free energies of protein-ligand complexes", Journal of Chemical Theory and Computation **6**, 3558–3568 (2010).

[4]S. J. Fox, J. Dziedzic, T. Fox, C. S. Tautermann, and C. K. Skylaris, "Density functional theory calculations on entire proteins for free energies of binding: Application to a model polar binding site", Proteins: Structure, Function and Bioinformatics **82**, 3335–3346 (2014).

[5]W. A. Baase, L. Liu, D. E. Tronrud, and B. W. Matthews, *Lessons from the lysozyme of phage T4*, Vol. 19, 4 (Wiley-Blackwell, Apr. 2010), pp. 631–641.