Electronic Supplementary Material (ESI) for Chemical Science.

SUPPORTING INFORMATION: Hybrid Classical/Machine-Learning Force Fields for the Accurate Description of Molecular Condensed-Phase Systems

Moritz Thürlemann, Sereina Riniker*

[*] *Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland. Email: sriniker@ethz.ch*

S1 Computational Cost

Figure S1 shows the time required for a single gradient call for a water box with 354 water molecules evaluated on an Intel Xeon E-2146G processor. Calculating intramolecular gradients and D3 gradients required 0.29 s and 1.49 s, respectively. Calculation of the intermolecular gradients with the ANA2B models was measured at 1.29 s (ANA2B^0), 1.74 s (ANA2B^1) and 13.82 s (ANA2B$^\infty$). For the D3 model, cutoffs were set to 15 Å, 8 Å, and 15 Å for the two-body-term, three-body-term, and the coordination number, respectively.

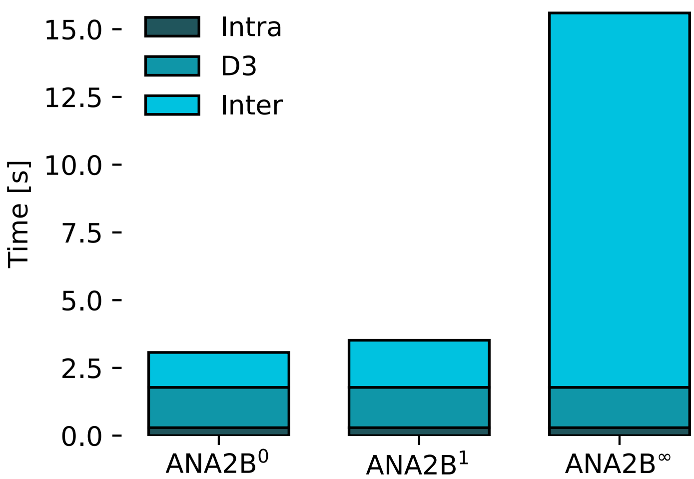

Figure S1: Computational cost for a single gradient evaluation using an Intel Xeon E-2146G processor.

S2 Intramolecular Training and Validation Sets

Table S1: Error statistics for the monomer training and validation sets in [kJ/mol] and [kJ/molA]: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Name	N	MAE	ME	RMSE	r^2
ΔEnergy Train	1'398'301	0.5	0.0	1.3	1.00
Gradients Train	110'948'431	0.8	0.0	18.1	0.99
ΔEnergy Validation	79'369	0.6	0.0	2.2	1.00
Gradients Validation	6'305'079	0.8	0.0	2.2	1.00

S3 Intramolecular Test Sets

Table S2: Error statistics for the monomer test sets in [kJ/mol]: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Name	N	MAE	ME	RMSE	r^2
Glucose	205	2.5	2.2	3.1	0.99
Maltose	223	2.7	1.5	3.3	0.97
SCONF	17	1.6	1.2	2.0	0.98
PCONF	10	6.7	6.7	6.9	0.92
ACONF	15	0.5	-0.5	0.6	1.00
CYCONF	10	2.7	-2.6	3.3	0.68

S4 Dimer Training Sets

Table S3: Error statistics for the dimer training sets in [kJ/mol]: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Model	Name	N	MAE	ME	RMSE	r^2
ANA2B^0	DES5M	4'034'267	1.9	0.1	5.4	1.00
ANA2B^1	DES5M	4'034'267	2.0	-0.7	5.7	1.00
ANA2B$^\infty$	DES5M	4'034'267	2.0	-0.7	5.7	1.00
ANA2B^0	DES370K	269'611	1.2	0.4	3.2	0.99
ANA2B^1	DES370K	269'611	1.2	-0.1	3.1	0.99
ANA2B$^\infty$	DES370K	269'611	1.1	0.2	2.9	1.00

S5 Dimer Validation Sets

Table S4: Error statistics for the dimer validation sets in [kJ/mol]: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Model	Name	N	MAE	ME	RMSE	r^2
ANA2B^0	S66x8	528	1.3	-0.6	2.5	0.99
ANA2B^1	S66x8	528	0.8	-0.3	1.3	1.00
ANA2B$^\infty$	S66x8	528	0.8	-0.4	1.2	1.00
ANA2B^0	S7L	7	21.1	19.7	29.9	0.93
ANA2B^1	S7L	7	2.0	1.9	3.5	1.00
ANA2B$^\infty$	S7L	7	2.3	-0.5	3.6	1.00

S6 Dimer Test Sets

Table S5: Error statistics for the dimer test sets in [kJ/mol] with the ANA2B^0 model: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Name	N	MAE	ME	RMSE	r^2
D1200	482	1.7	0.7	2.7	0.91
D442x10	1570	1.9	0.6	3.8	0.88
R739x5	1615	2.5	0.5	3.7	0.87
HB300SPXx10	1210	4.1	-3.7	7.6	0.88
HB375x10	3750	1.9	-1.2	3.6	0.96
ACHC	54	4.8	4.7	5.0	0.96
BBI	100	1.0	-0.9	1.1	1.00
HBC1	58	8.5	-8.4	14.2	0.94
HSG	16	0.8	-0.1	0.9	1.00
JSCH	123	4.7	-1.5	7.1	0.99
S22	22	3.3	-1.7	4.6	0.99
S22x7	154	5.9	-3.2	12.1	0.98
SSI	2596	0.6	0.2	1.0	0.99
UBQ	81	1.0	0.3	2.5	0.99

Table S6: Error statistics for the dimer test sets in [kJ/mol] with the $ANA2B^1$ model: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Name	N	MAE	ME	RMSE	r^2
D1200	482	1.2	-0.1	1.9	0.94
D442x10	1570	1.6	-0.0	3.5	0.89
R739x5	1615	2.3	0.6	3.3	0.89
HB300SPXx10	1210	3.1	-2.6	5.8	0.93
HB375x10	3750	1.4	-0.4	2.4	0.98
ACHC	54	2.2	2.2	2.6	0.96
BBI	100	0.7	0.1	1.0	1.00
HBC1	58	3.3	-3.2	5.8	0.99
HSG	16	0.7	0.2	0.9	1.00
JSCH	123	2.9	-1.8	4.3	1.00
S22	22	1.7	-1.1	2.5	1.00
S22x7	154	3.0	-0.4	6.8	0.99
SSI	2596	0.7	0.4	1.1	0.99
UBQ	81	1.2	0.9	2.6	0.99

Table S7: Error statistics for the dimer test sets in [kJ/mol] with the $ANA2B^\infty$ model: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). For gradients, N refers to the number of gradient components.

Name	N	MAE	ME	RMSE	r^2
D1200	482	1.2	-0.3	1.9	0.94
D442x10	1570	1.5	-0.2	3.2	0.91
R739x5	1615	2.3	0.3	3.2	0.90
HB300SPXx10	1210	3.5	-3.1	6.5	0.91
HB375x10	3750	1.4	-0.7	2.3	0.98
ACHC	54	1.0	0.3	1.3	0.97
BBI	100	0.7	-0.5	0.8	1.00
HBC1	58	2.0	-0.4	3.6	0.99
HSG	16	0.8	-0.1	0.9	1.00
JSCH	123	2.8	-2.0	4.1	1.00
S22	22	1.6	-1.0	2.2	1.00
S22x7	154	2.8	-0.4	6.3	0.99
SSI	2596	0.6	0.2	1.0	0.99
UBQ	81	1.0	0.6	2.5	0.99

S7 Molecular Crystals Test Sets

S7.1 ANA2B^0

Table S8: Reference and calculated lattice energies with ANA2B^0 for the X23 and ICE13 sets in [kJ/mol]. Δ refers to the difference (pred. - exp.).

Dataset	Name	Exp.	Pred.	Δ
X23	CYTSIN01	-163.5	-154.0	9.5
	UREAXX12	-102.1	-104.1	-2.0
	HXMTAM10	-84.1	-96.5	-12.4
	PYRZOL05	-78.8	-78.4	0.4
	CYHEXO	-90.0	-93.8	-3.8
	ACETAC07	-73.6	-65.9	7.7
	ADAMAN08	-71.8	-64.5	7.3
	ANTCEN09	-110.4	-114.3	-3.9
	BENZEN01	-54.8	-54.0	0.8
	CYANAM01	-81.5	-74.8	6.7
	ECARBM01	-88.2	-85.3	2.9
	FORMAM02	-81.1	-75.9	5.2
	IMAZOL04	-90.4	-83.1	7.3
	NAPHTA23	-81.3	-84.7	-3.4
	OXALAC03	-98.8	-91.5	7.3
	OXALAC04	-96.8	-98.5	-1.7
	PYRAZI01	-64.3	-67.6	-3.3
	TRIZIN	-62.6	-71.9	-9.3
	TROXAN11	-64.6	-60.5	4.1
	SUCACB03	-130.1	-131.6	-1.5
	URACIL	-136.2	-133.0	3.2
	NH3	-38.7	-37.9	0.8
	CO2	-29.4	-31.7	-2.3
ICE13	Ih	-59.4	-46.3	13.1
	II	-59.1	-48.9	10.2
	III	-58.2	-48.7	9.5
	IV	-55.6	-50.5	5.1
	VI	-57.7	-50.4	7.3
	VII	-54.5	-50.7	3.8
	VIII	-55.2	-52.0	3.2
	IX	-58.8	-49.6	9.2
	XI	-58.3	-47.7	10.6
	XIII	-57.3	-49.9	7.4
	XIV	-57.8	-50.5	7.3
	XV	-57.7	-50.5	7.2
	XVII	-57.7	-46.3	11.4

Table S9: Reference and calculated lattice energies with ANA2B[1] for the X23 and ICE13 sets in [kJ/mol]. Δ refers to the difference (pred. - exp.).

Dataset	Name	Exp.	Pred.	Δ
X23	CYTSIN01	-163.5	-157.8	5.7
	UREAXX12	-102.1	-104.5	-2.4
	HXMTAM10	-84.1	-90.8	-6.7
	PYRZOL05	-78.8	-78.5	0.3
	CYHEXO	-90.0	-95.5	-5.5
	ACETAC07	-73.6	-70.0	3.6
	ADAMAN08	-71.8	-67.5	4.3
	ANTCEN09	-110.4	-111.3	-0.9
	BENZEN01	-54.8	-55.6	-0.8
	CYANAM01	-81.5	-81.1	0.4
	ECARBM01	-88.2	-85.4	2.8
	FORMAM02	-81.1	-75.0	6.1
	IMAZOL04	-90.4	-87.7	2.7
	NAPHTA23	-81.3	-83.3	-2.0
	OXALAC03	-98.8	-93.7	5.1
	OXALAC04	-96.8	-97.5	-0.7
	PYRAZI01	-64.3	-66.3	-2.0
	TRIZIN	-62.6	-66.3	-3.7
	TROXAN11	-64.6	-62.0	2.6
	SUCACB03	-130.1	-134.6	-4.5
	URACIL	-136.2	-129.8	6.4
	NH3	-38.7	-40.4	-1.7
	CO2	-29.4	-28.9	0.5
ICE13	Ih	-59.4	-56.2	3.2
	II	-59.1	-58.3	0.8
	III	-58.2	-55.6	2.6
	IV	-55.6	-56.4	-0.8
	VI	-57.7	-57.9	-0.2
	VII	-54.5	-53.8	0.7
	VIII	-55.2	-59.2	-4.0
	IX	-58.8	-57.5	1.3
	XI	-58.3	-57.1	1.2
	XIII	-57.3	-58.5	-1.2
	XIV	-57.8	-57.5	0.3
	XV	-57.7	-57.9	-0.2
	XVII	-57.7	-55.3	2.4

Table S10: Reference and calculated lattice energies with ANA2B$^{\infty}$ for the X23 and ICE13 sets in [kJ/mol]. Δ refers to the difference (pred. - exp.).

Dataset	Name	Exp.	Pred.	Δ
X23	CYTSIN01	-163.5	-163.3	0.2
	UREAXX12	-102.1	-107.8	-5.7
	HXMTAM10	-84.1	-92.1	-8.0
	PYRZOL05	-78.8	-80.0	-1.2
	CYHEXO	-90.0	-93.3	-3.3
	ACETAC07	-73.6	-67.7	5.9
	ADAMAN08	-71.8	-66.2	5.6
	ANTCEN09	-110.4	-112.4	-2.0
	BENZEN01	-54.8	-55.0	-0.2
	CYANAM01	-81.5	-82.7	-1.2
	ECARBM01	-88.2	-85.5	2.7
	FORMAM02	-81.1	-77.7	3.4
	IMAZOL04	-90.4	-90.7	-0.3
	NAPHTA23	-81.3	-83.5	-2.2
	OXALAC03	-98.8	-95.6	3.2
	OXALAC04	-96.8	-99.6	-2.8
	PYRAZI01	-64.3	-64.5	-0.2
	TRIZIN	-62.6	-64.3	-1.7
	TROXAN11	-64.6	-61.5	3.1
	SUCACB03	-130.1	-134.8	-4.7
	URACIL	-136.2	-132.0	4.2
	NH3	-38.7	-42.0	-3.3
	CO2	-29.4	-31.7	-2.3
ICE13	Ih	-59.4	-60.2	-0.8
	II	-59.1	-61.5	-2.4
	III	-58.2	-58.1	0.1
	IV	-55.6	-56.1	-0.5
	VI	-57.7	-59.0	-1.3
	VII	-54.5	-55.6	-1.1
	VIII	-55.2	-54.7	0.5
	IX	-58.8	-60.9	-2.1
	XI	-58.3	-60.3	-2.0
	XIII	-57.3	-59.9	-2.6
	XIV	-57.8	-58.3	-0.5
	XV	-57.7	-58.8	-1.1
	XVII	-57.7	-59.7	-2.0

S8 Molecular Liquids

Table S11: Error statistics for properties of molecular liquids in the GROMOS 2016H66 set using the ANA2B^{1} model: mean absolute error (MAE), mean error (ME), root-mean-square error (RMSE), and correlation coefficient (r^2). H_{vap} is given in [kJ/mol] and ρ in [kg·m^{-3}]. N refers to the number of liquids.

Name	N	MAE	ME	RMSE	r^2
H_{vap} [kJ/mol]	57	2.2 ± 0.9	-0.6	2.8	0.97
Density [kg· m^{-3}]	57	28.4 ± 5.8	-25.4	33.9	0.97

Table S12: Experimental and predicted properties for the molecular liquids in the GROMOS 2016H66 set using the ANA2B[1] model. Standard deviations were obtained over three replica. H_{vap} is given in [kJ/mol], ρ in [kg·m^{-3}] and T in [K].

Name	H_{vap} exp.	H_{vap} sim.	ρ exp.	ρ sim.	T exp.	T sim.
MTL	37.43	37.14±0.37	784.00	791.79±7.22	298.15	300.12
ETL	42.31	41.61±0.49	784.93	803.76±6.04	298.15	299.98
1PL	47.49	45.07±0.57	799.60	814.49±4.84	298.15	300.07
BTL	52.34	51.07±2.05	805.75	831.02±3.25	298.15	300.19
PTL	56.94	54.91±0.97	810.80	833.86±8.60	298.15	300.25
HXL	61.85	57.88±1.71	815.34	832.88±0.65	298.15	299.83
HPL	66.81	65.53±3.19	820.00	830.84±5.63	298.15	299.76
OTL	70.98	66.77±3.44	821.57	838.01±4.11	298.15	299.98
2PL	45.52	44.78±0.30	780.00	802.24±8.97	298.15	300.16
2BTL	49.66	48.88±1.74	802.41	803.07±8.52	298.15	300.00
2PTL	53.10	51.76±2.68	805.40	820.88±2.53	298.15	299.88
3PTL	53.10	52.75±0.07	816.00	820.17±4.56	298.15	299.85
CHXL	62.01	56.83±0.75	968.40	945.38±8.17	298.15	299.88
2M2P	46.82	47.22±0.85	781.20	821.72±11.20	298.15	300.03
2M2B	50.20	49.22±0.48	805.00	835.20±5.89	298.15	300.05
DME	21.24	20.04±0.11	722.00	715.95±13.12	254.00	255.30
DEE	27.10	26.10±1.12	707.82	704.38±11.72	298.15	299.95
MPH	27.60	27.09±2.25	735.60	714.69±17.36	298.15	299.95
DXE	36.39	36.04±1.46	863.70	873.24±6.89	298.15	299.63
EAL	26.11	28.38±0.43	778.00	819.67±5.34	298.15	299.48
PAL	29.63	32.16±1.89	791.20	821.19±4.30	298.15	299.83
BAL	33.68	35.49±0.82	796.40	822.37±1.87	298.15	299.96
PPN	31.30	34.26±0.34	784.40	830.88±3.58	298.15	299.68
BTN	34.51	37.08±1.22	799.70	831.34±6.31	298.15	299.48
2PN	38.40	40.92±1.63	801.50	839.02±2.34	298.15	299.82
3PN	38.56	40.20±1.13	809.45	834.60±8.29	298.15	299.72
2HN	42.90	48.17±0.97	806.70	851.21±1.46	298.15	299.80
3HN	42.47	38.36±1.60	810.00	832.17±5.00	298.15	299.78
ACA	51.60	53.30±0.17	1043.92	1118.45±4.21	298.15	299.45
PPA	57.30	56.60±1.49	988.08	1049.95±4.21	298.15	299.74
BTA	58.00	59.00±0.85	953.20	1000.59±4.13	298.15	299.99
EAE	35.62	37.70±0.67	894.55	938.99±3.68	298.15	299.53
MPE	35.85	35.97±0.80	909.00	934.95±5.19	298.15	299.46
PAE	39.83	40.90±0.62	883.03	918.74±1.82	298.15	299.53
BAE	43.64	45.10±0.70	876.36	905.53±7.92	298.15	299.68
EAN	26.60	26.36±0.18	677.00	680.39±2.24	298.15	299.72
PAN	31.26	30.81±0.57	712.10	708.21±16.82	298.15	299.89
BAN	35.74	34.23±0.58	736.83	732.85±2.42	298.15	299.82
DEAN	31.32	30.52±0.93	701.60	692.01±11.65	298.15	300.00
TEAN	34.81	30.24±0.65	723.05	712.41±5.91	298.15	299.70
EDAN	46.00	47.40±0.18	893.10	909.08±1.31	298.15	300.15
AMD	56.10	59.49±1.02	989.20	1013.44±3.69	364.00	365.82
LNMA	54.39	56.61±0.15	926.00	931.87±18.82	333.00	334.58
DAMD	49.15	46.04±0.43	936.34	931.93±5.04	298.15	299.80
ETSH	27.30	29.99±0.51	833.00	885.94±4.23	298.15	299.55
PRSH	31.89	34.47±0.48	841.10	886.27±4.60	298.15	299.55
BTSH	36.53	40.14±0.59	836.74	881.69±4.78	298.15	300.03
DMS	27.65	28.27±0.45	842.28	863.42±0.74	298.15	299.45

DES	35.77	37.72±1.36	831.18	859.50±9.36	298.15	299.37
EMS	31.85	32.54±0.62	832.00	865.21±3.43	298.15	299.62
DMDS	38.33	41.74±0.32	1062.50	1114.99±2.88	298.15	299.39
BZN	33.84	40.94±0.12	873.60	941.13±1.82	298.15	299.33
TOL	37.99	44.58±0.31	862.19	921.38±2.82	298.15	299.44
PHL	55.40	60.16±0.60	1054.46	1095.06±9.52	333.00	334.54
PHT	64.96	64.90±0.92	1018.50	1061.54±1.14	309.00	310.64
ANLN	55.83	59.89±0.38	1017.50	1064.77±3.89	298.15	299.46
PYRI	40.41	46.28±0.68	978.24	1043.37±3.97	298.15	299.47

S9 Convergence of MD Simulations

Figures S4-S54 show the calculated density, temperature, and potential energy per molecule as a function of the simulation length. Time averages were taken over the interval ranging from 0 to 35 ps.

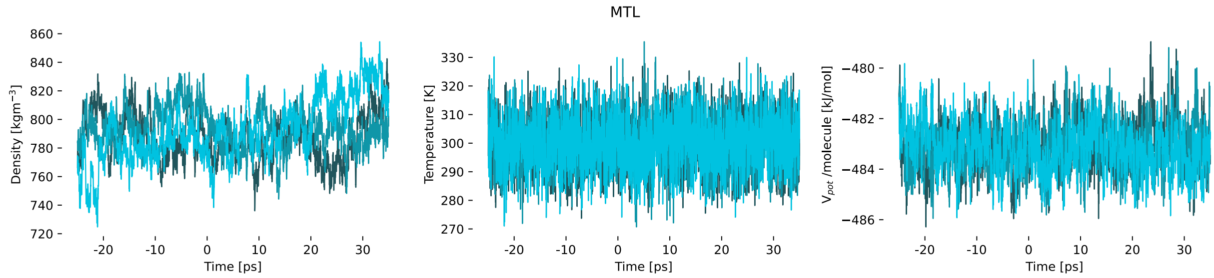

Figure S2: MTL

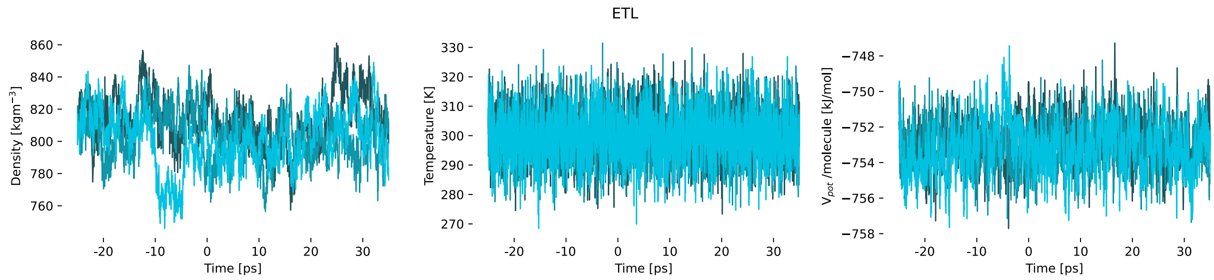

Figure S3: ETL

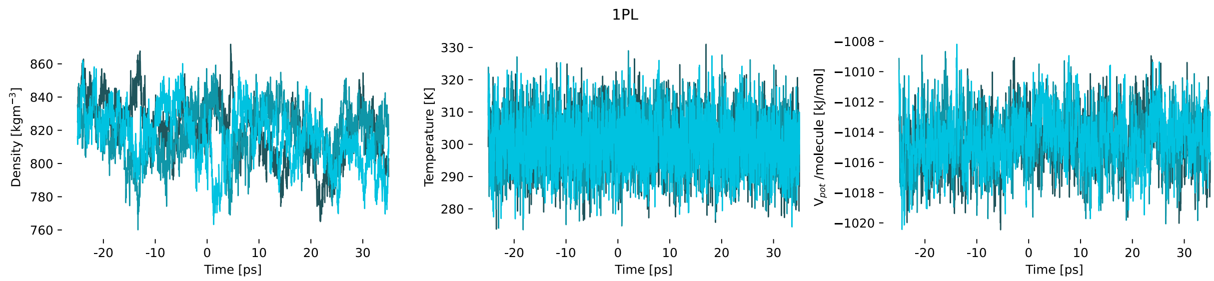

Figure S4: 1PL

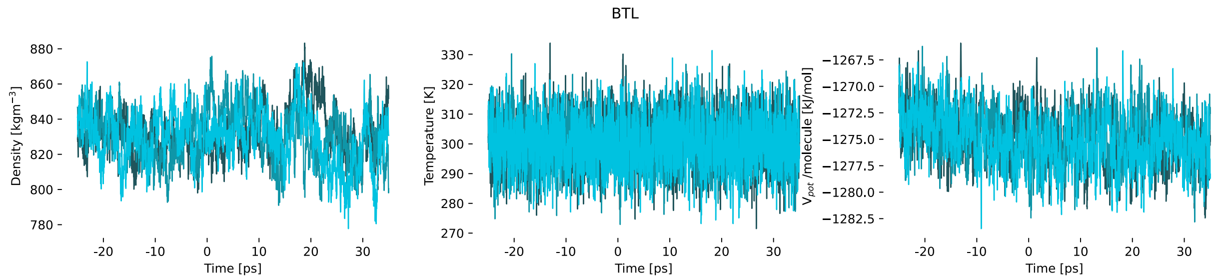

Figure S5: BTL

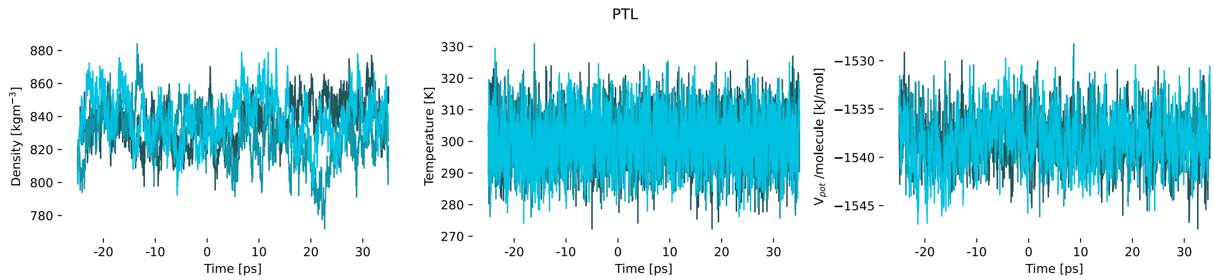

Figure S6: PTL

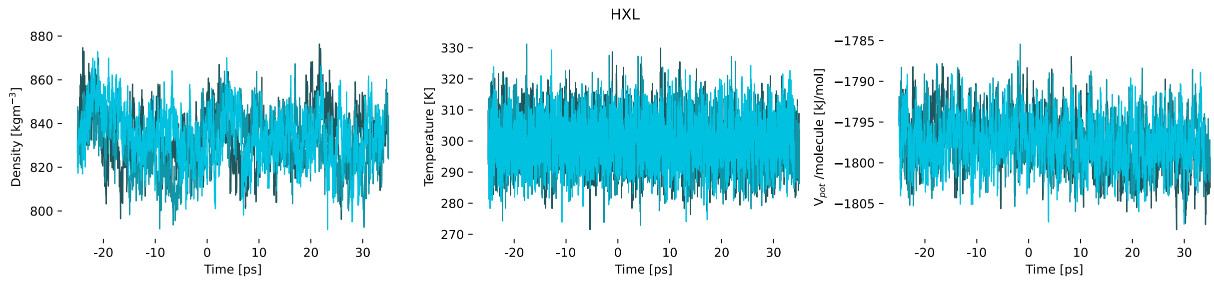

Figure S7: HXL

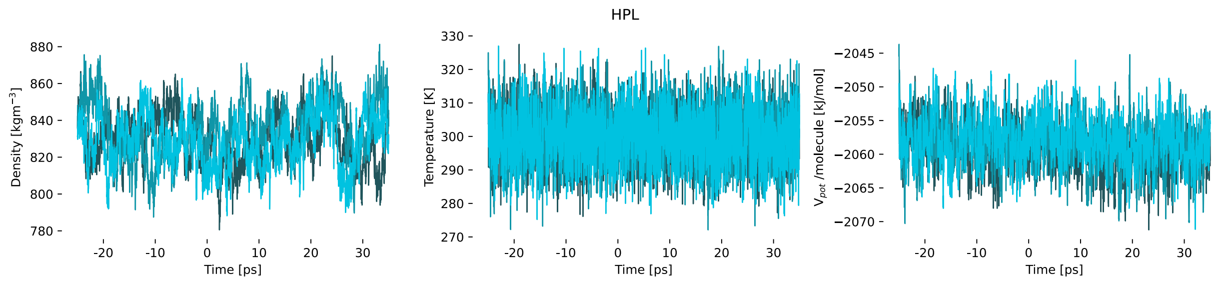

Figure S8: HPL

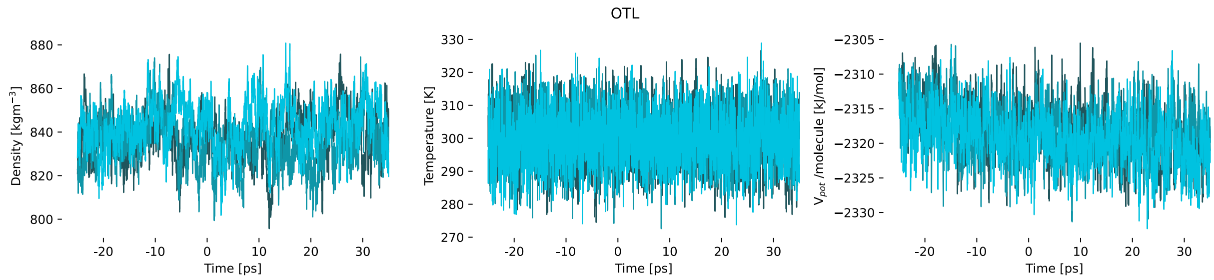

Figure S9: OTL

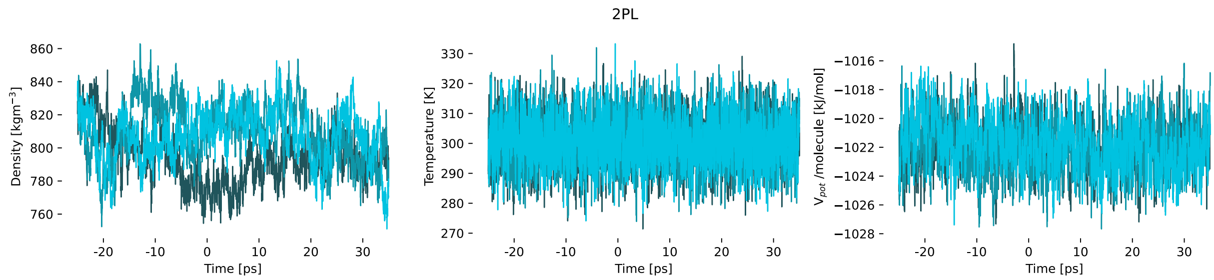

Figure S10: 2PL

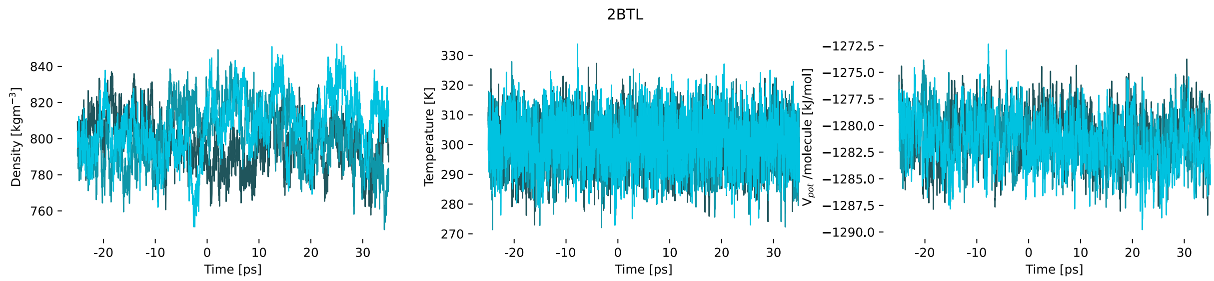

Figure S11: 2BTL

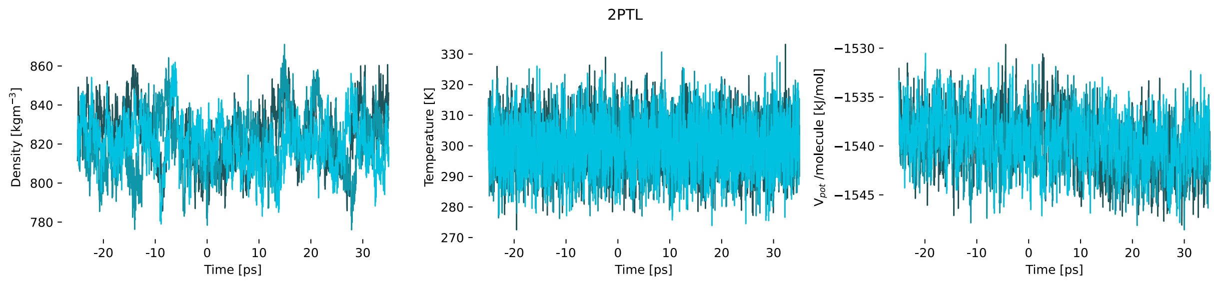

Figure S12: 2PTL

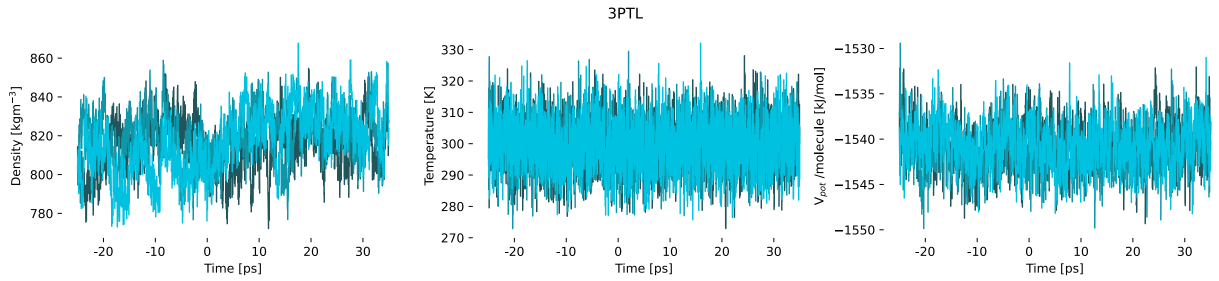

Figure S13: 3PTL

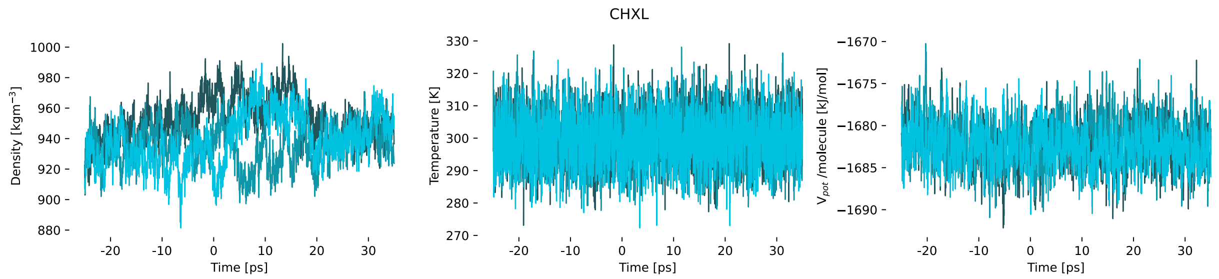

Figure S14: CHXL

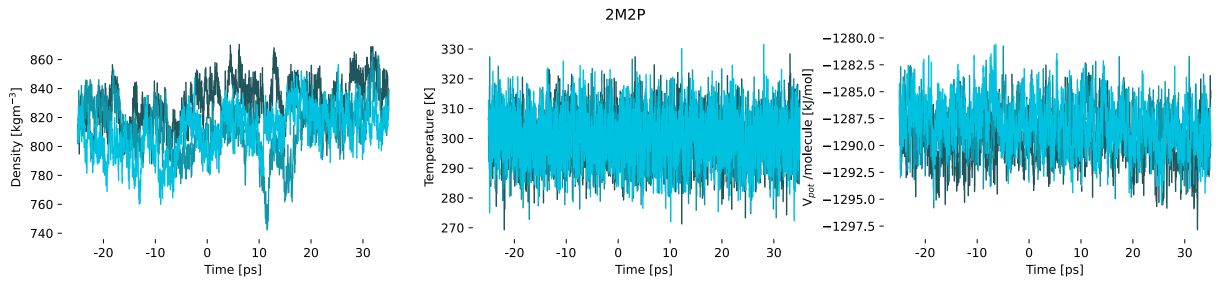

Figure S15: 2M2P

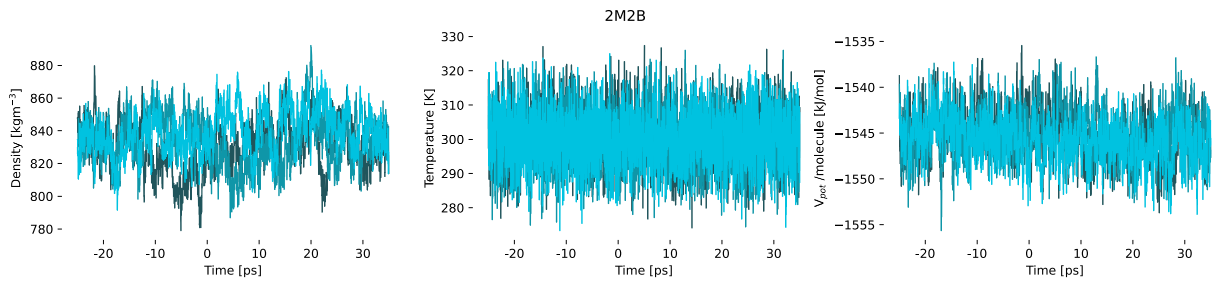

Figure S16: 2M2B

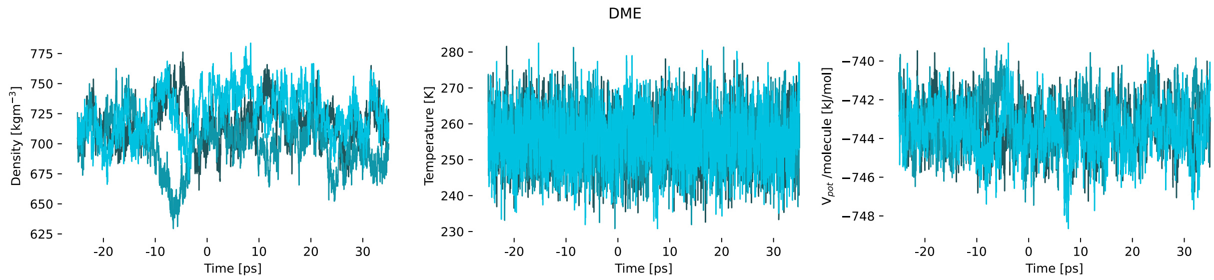

Figure S17: DME

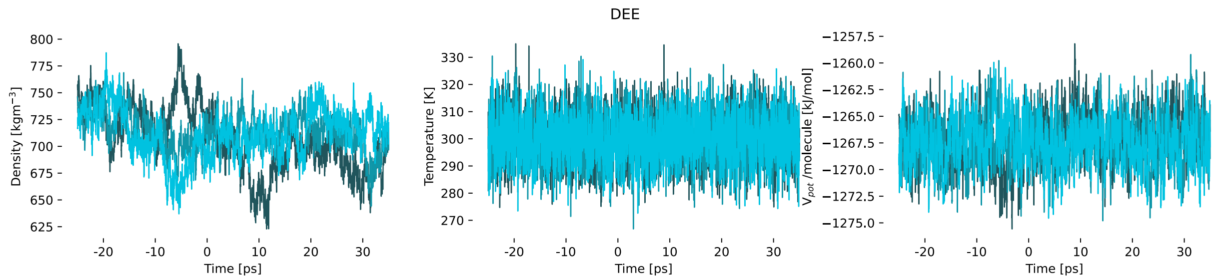

Figure S18: DEE

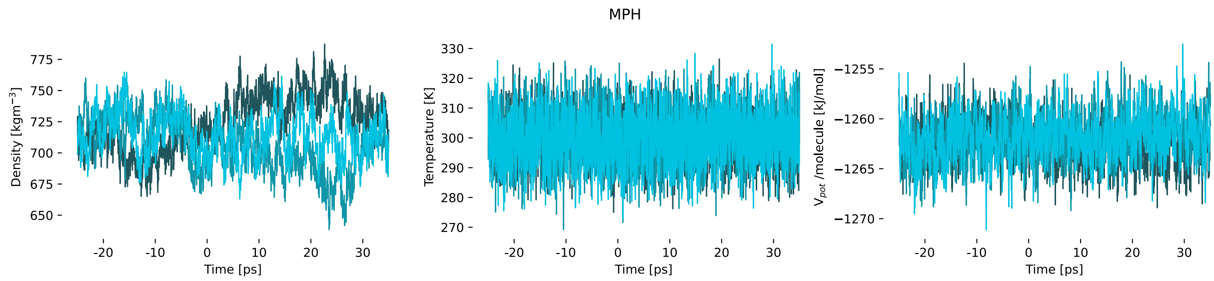

Figure S19: MPH

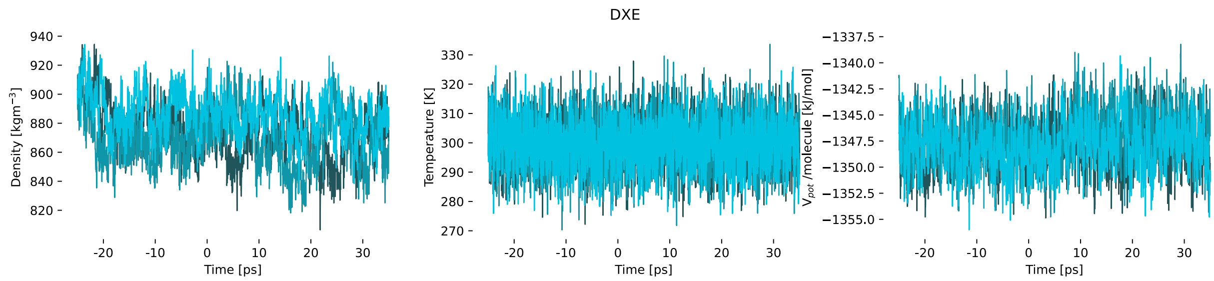

Figure S20: DXE

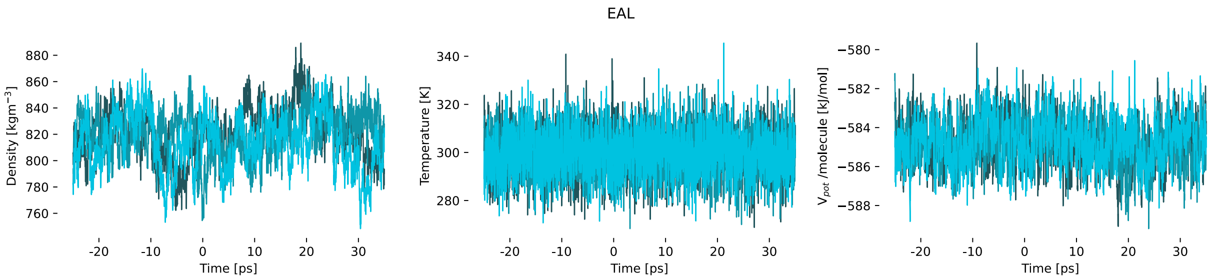

Figure S21: EAL

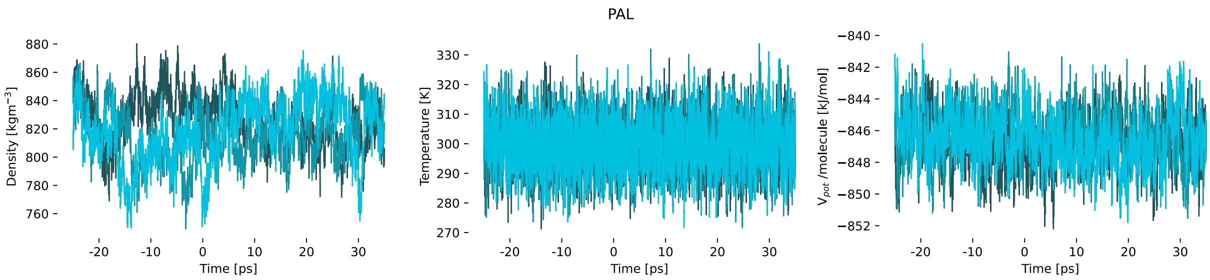

Figure S22: PAL

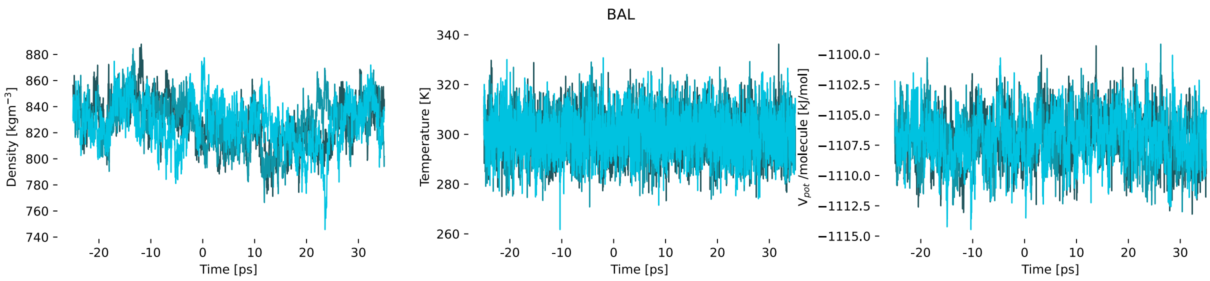

Figure S23: BAL

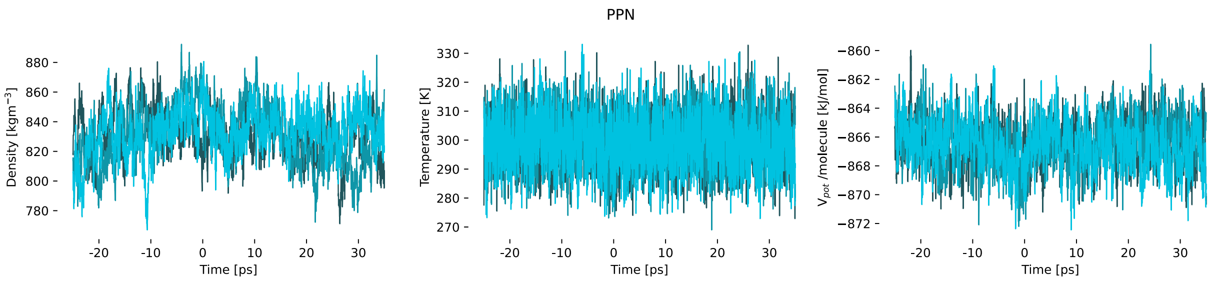

Figure S24: PPN

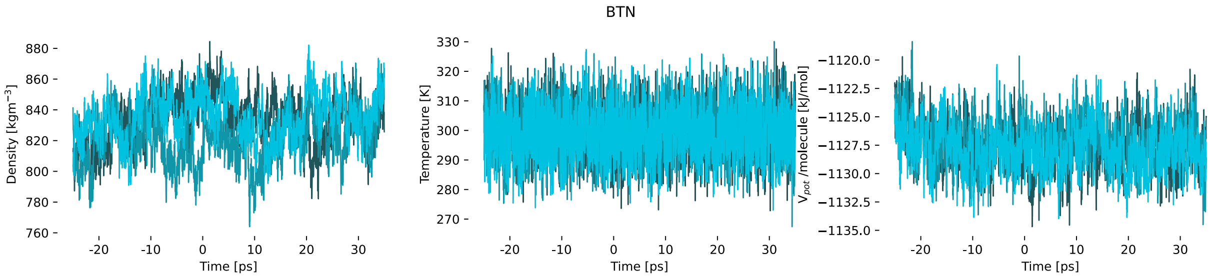

Figure S25: BTN

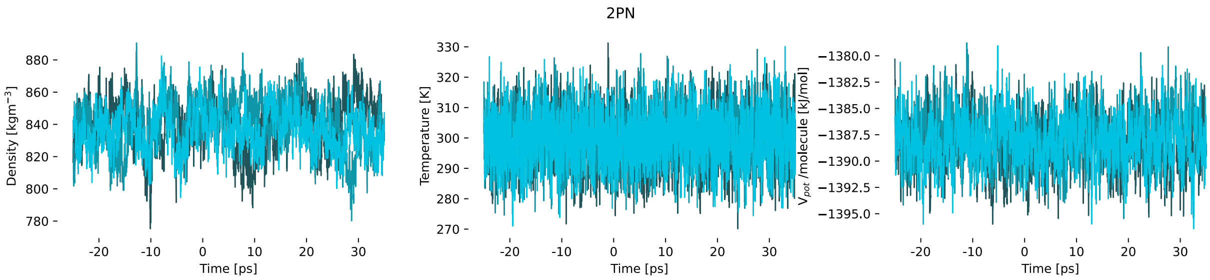

Figure S26: 2PN

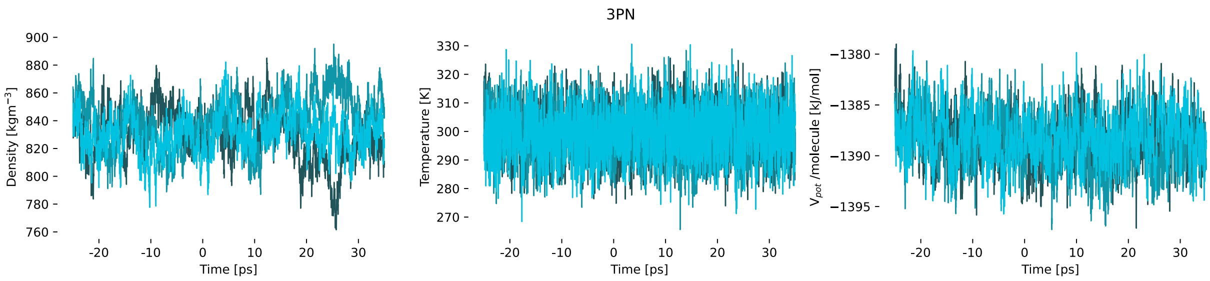

Figure S27: 3PN

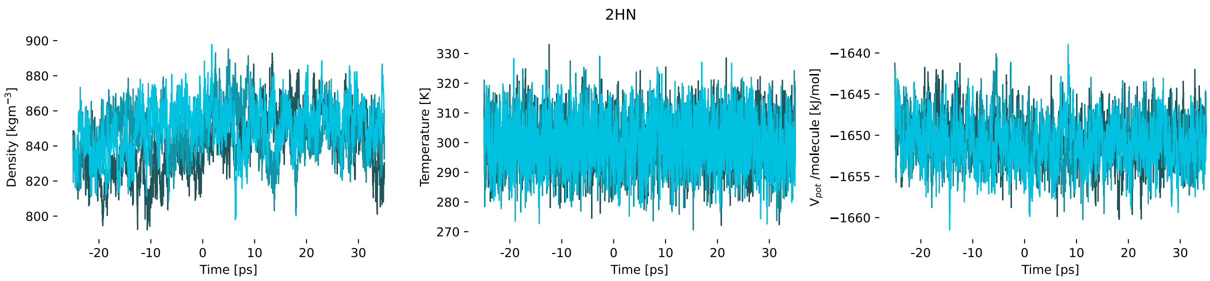

Figure S28: 2HN

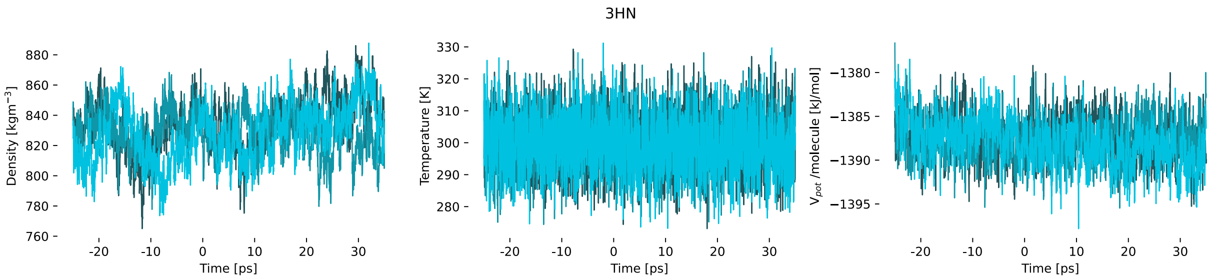

Figure S29: 3HN

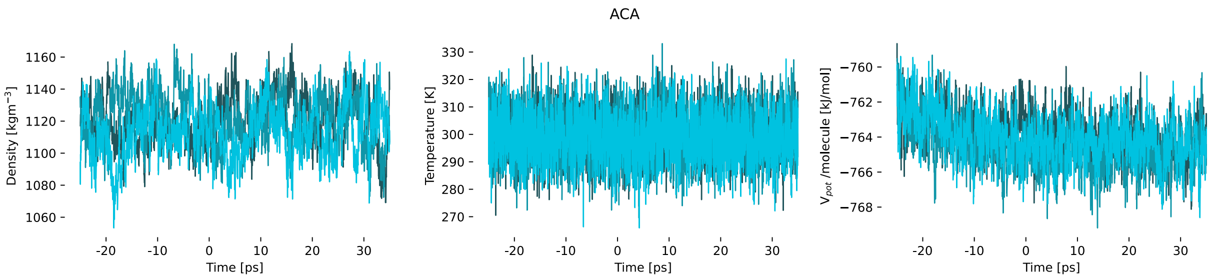

Figure S30: ACA

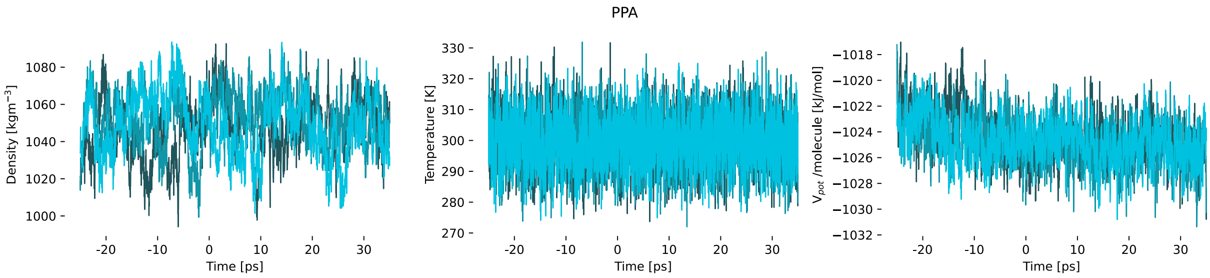

Figure S31: PPA

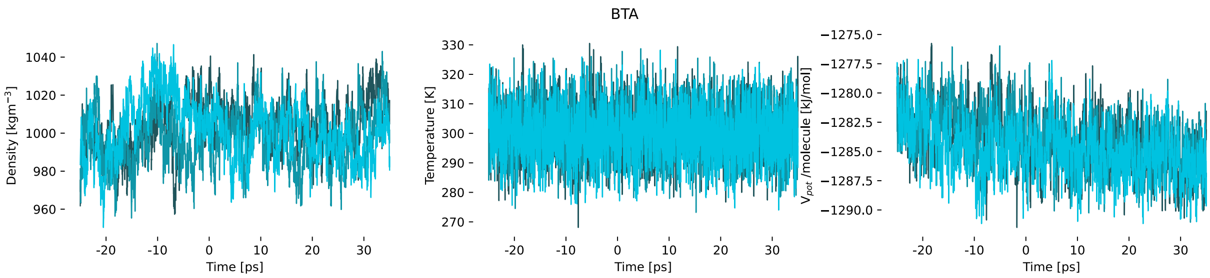

Figure S32: BTA

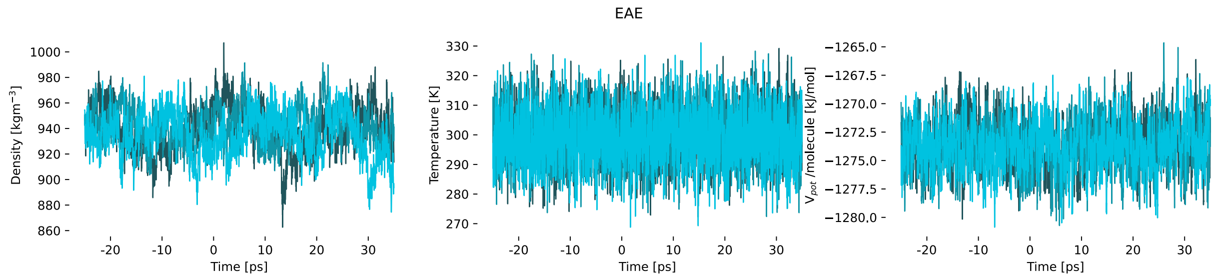

Figure S33: EAE

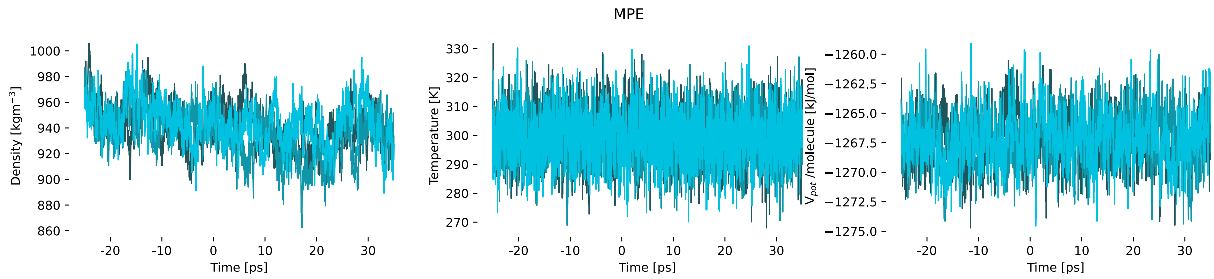

Figure S34: MPE

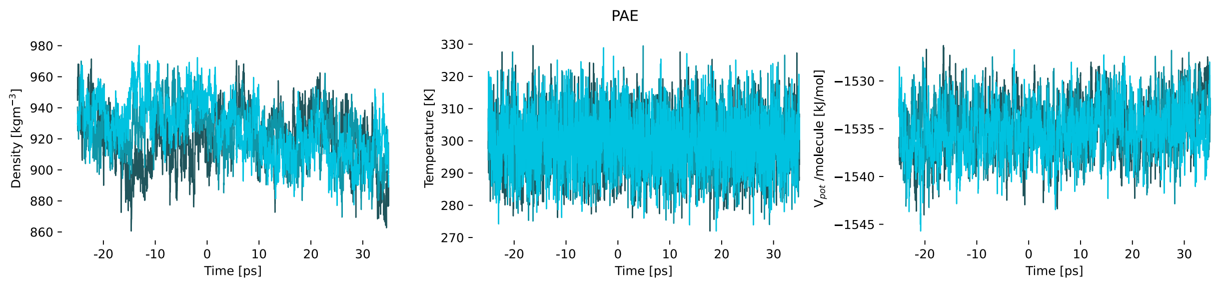

Figure S35: PAE

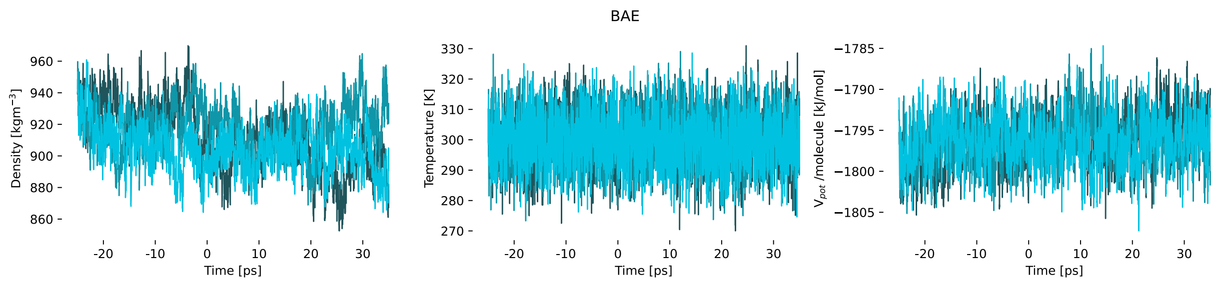

Figure S36: BAE

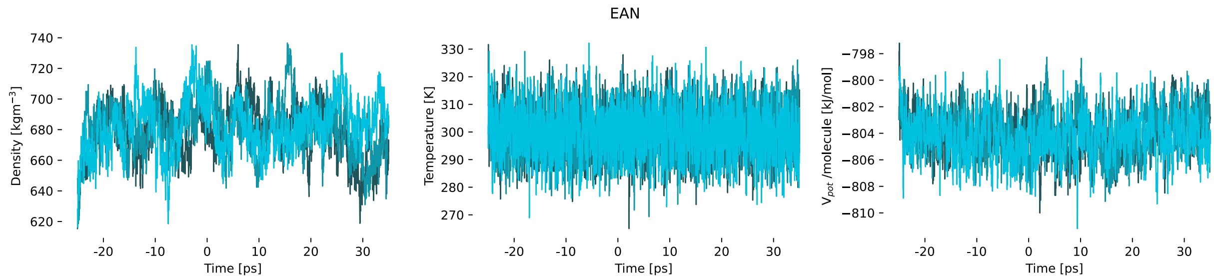

Figure S37: EAN

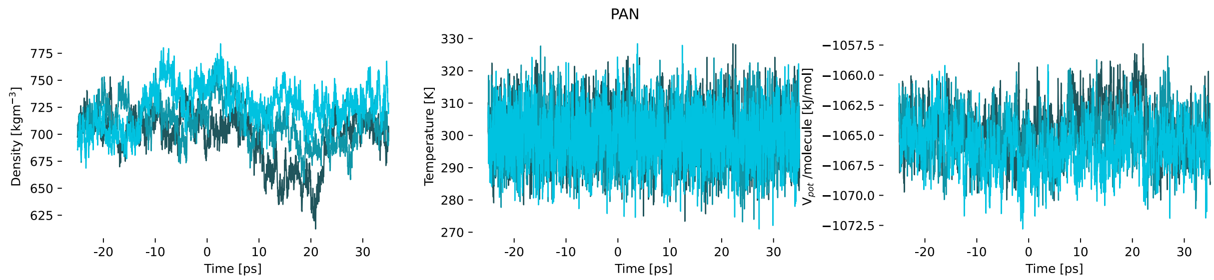

Figure S38: PAN

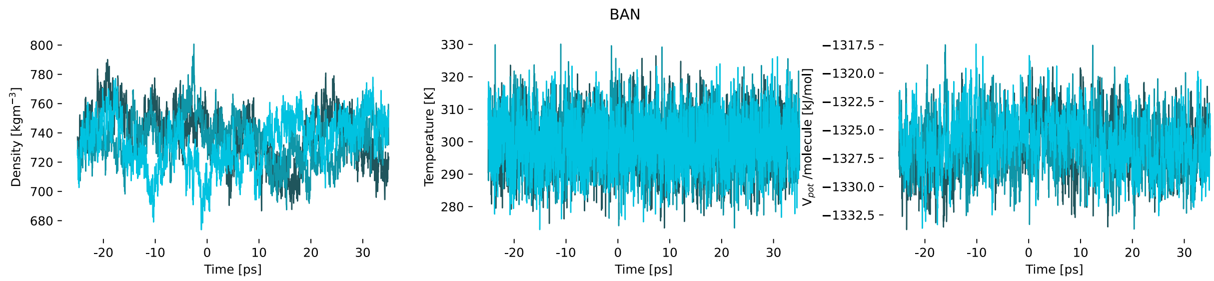

Figure S39: BAN

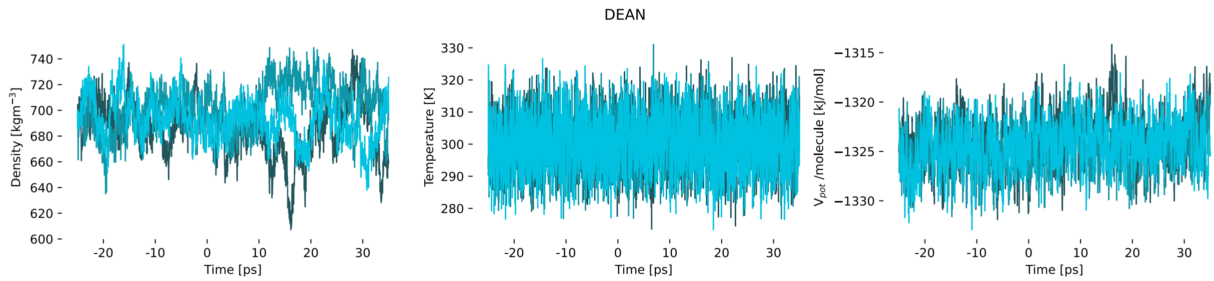

Figure S40: DEAN

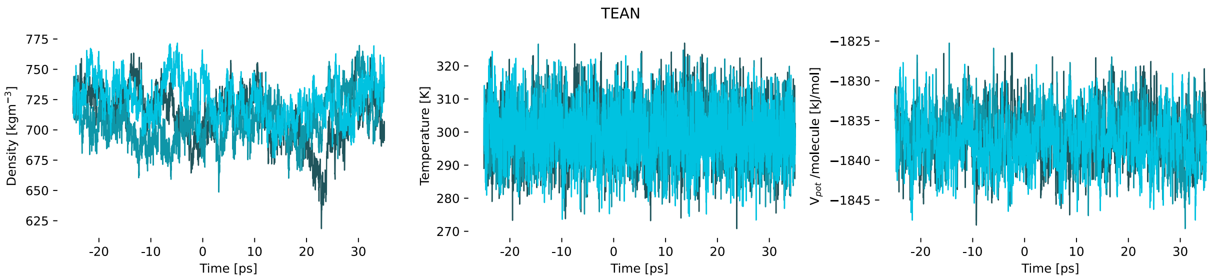

Figure S41: TEAN

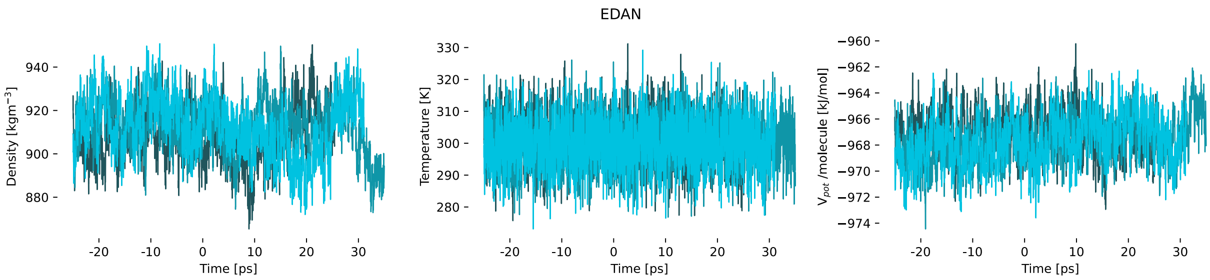

Figure S42: EDAN

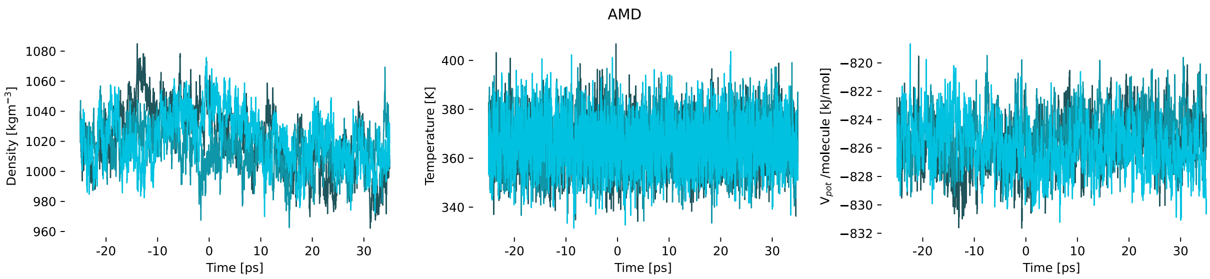

Figure S43: AMD

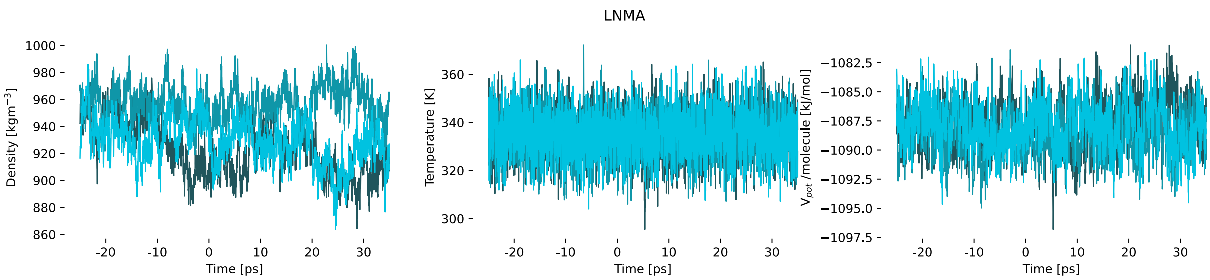

Figure S44: LNMA

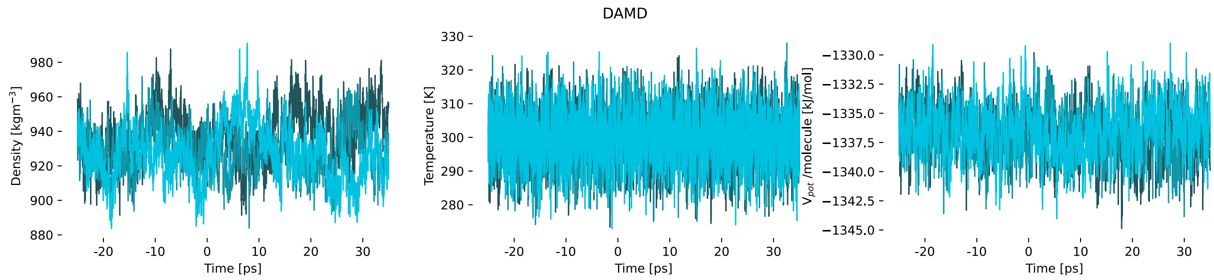

Figure S45: DAMD

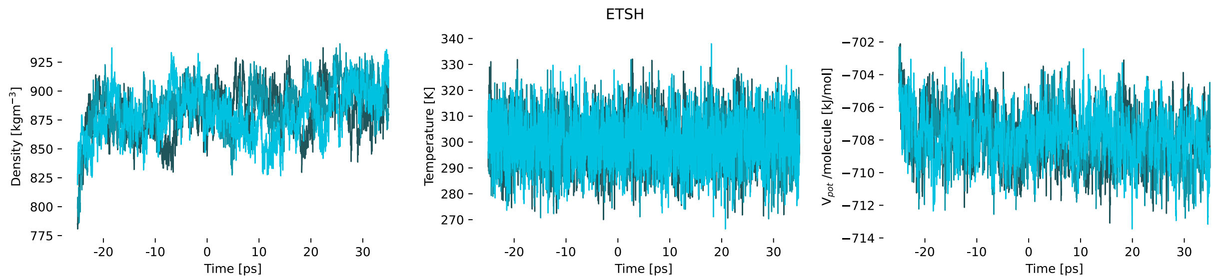

Figure S46: ETSH

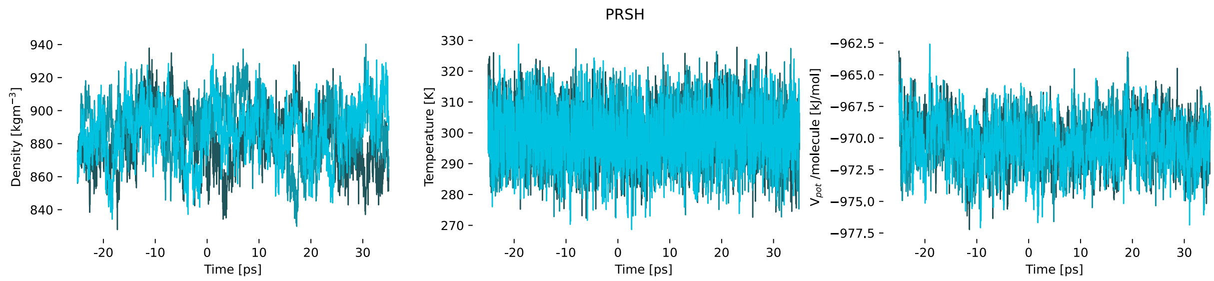

Figure S47: PRSH

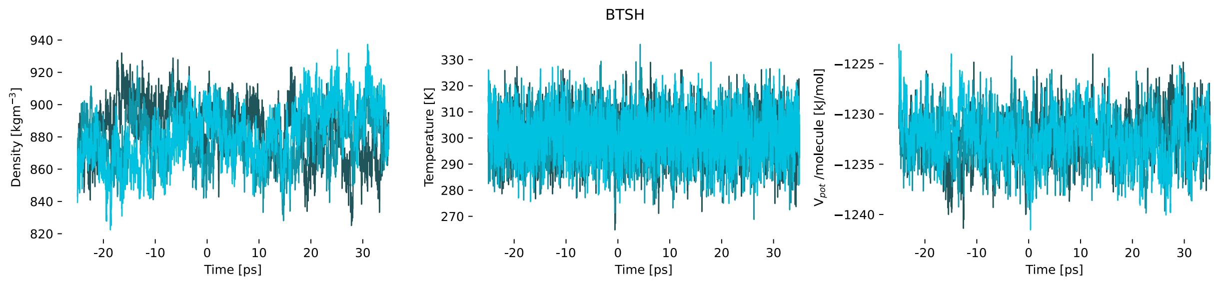

Figure S48: BTSH

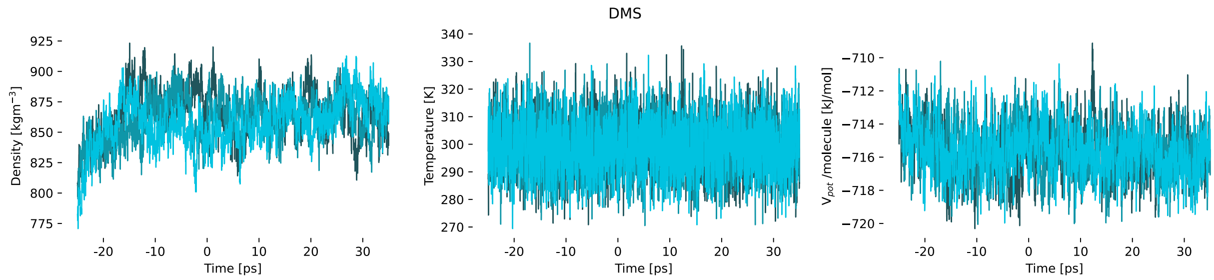

Figure S49: DMS

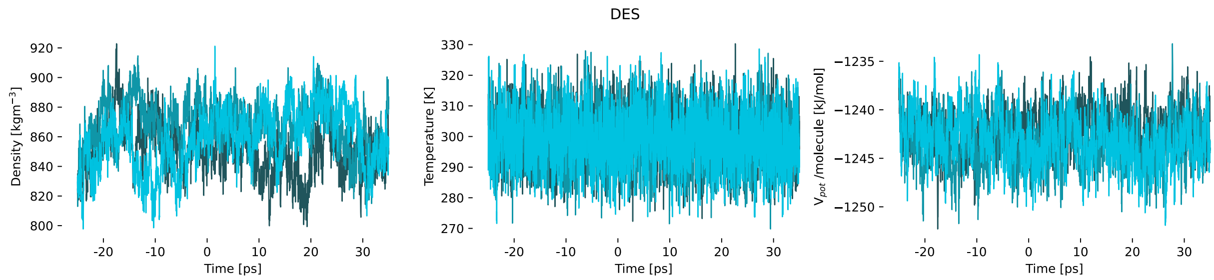

Figure S50: DES

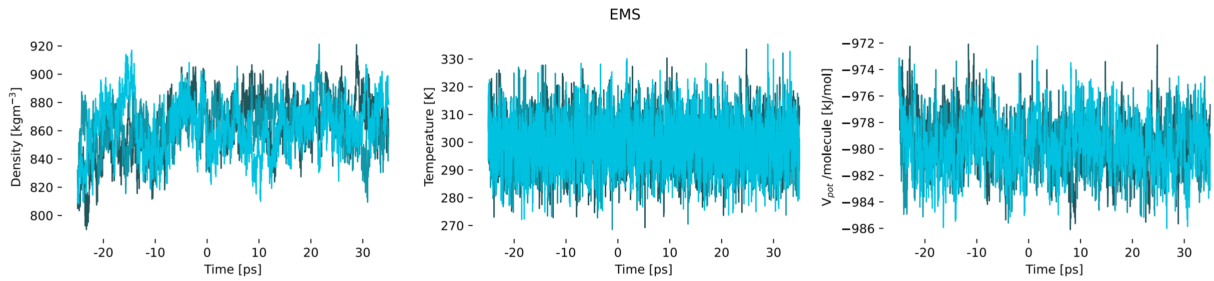

Figure S51: EMS

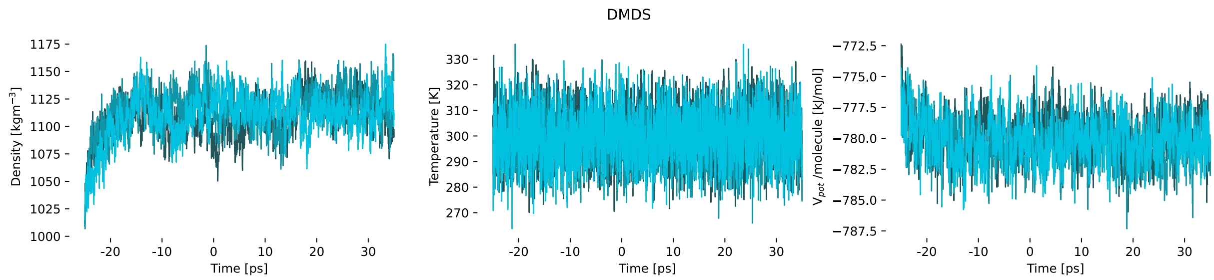

Figure S52: DMDS

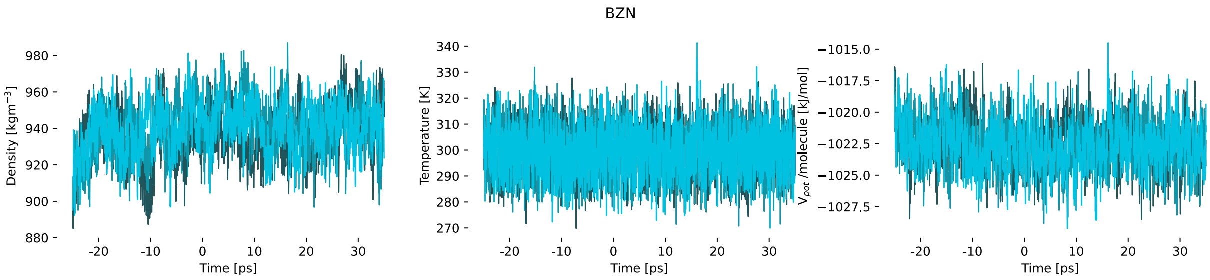

Figure S53: BZN

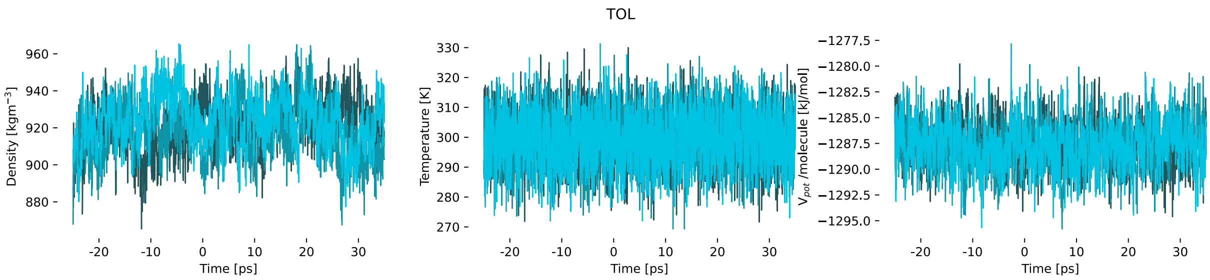

Figure S54: TOL

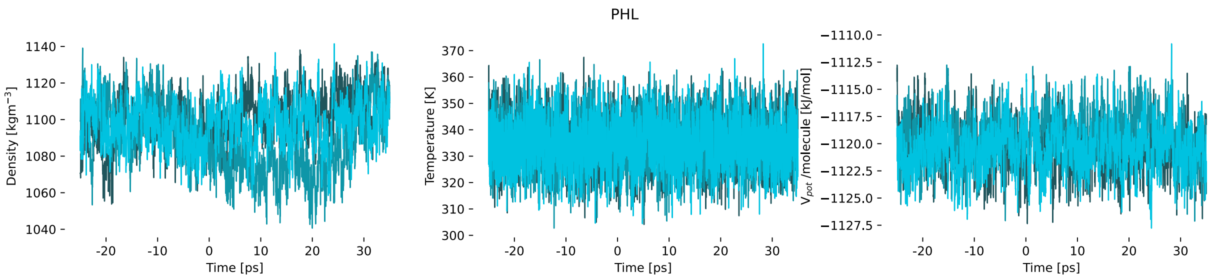

Figure S55: PHL

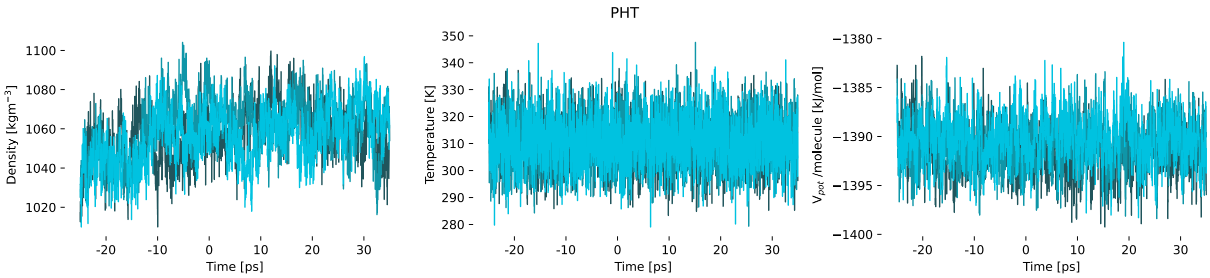

Figure S56: PHT

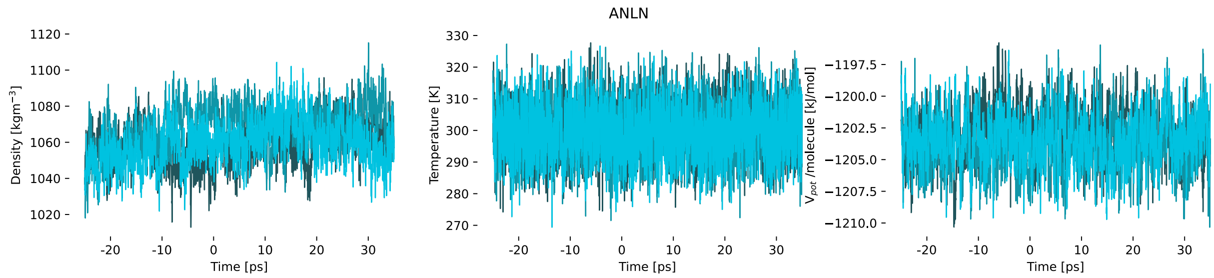

Figure S57: ANLN

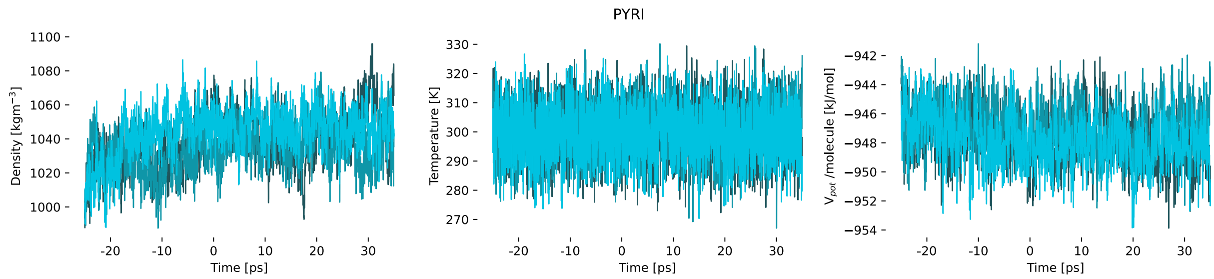

Figure S58: PYRI