Supporting Information: Configuration-Sampling-Based Surrogate Models for Rapid Parameterization of Non-bonded Interactions

Richard A. Messerly,[*,†] S. Mostafa Razavi,[‡] and Michael R. Shirts[¶]

†*Thermodynamics Research Center, National Institute of Standards and Technology, Boulder, Colorado, 80305*

‡*Department of Chemical and Biomolecular Engineering, The University of Akron, Akron, Ohio, 44325*

¶*Department of Chemical and Biological Engineering, Boulder, Colorado, 80309*

E-mail: richard.messerly@nist.gov

SI.I Simulation Set-Up

This section is provided to improve the reproducibility of the results presented in this study.

SI.I.1 State Points

Tables SI.I, SI.II, SI.III, SI.IV, and SI.V contain the state points that were simulated for ethane, hexafluoroethane, propane, *n*-butane, and *n*-octane, respectively. The first 10 state points of each table correspond to five isochores while the last 9 points are for the supercritical isotherm. The number of state points, the specified reduced temperatures, and the spacing between neighboring densities were recommended by the developers of the ITIC approach (J. Richard Elliott and Seyed Mostafa Razavi). It has been demonstrated that these points are sufficient for accurately calculating $\rho_{\mathrm{l}}^{\mathrm{sat}}$, $\rho_{\mathrm{v}}^{\mathrm{sat}}$, and $P_{\mathrm{v}}^{\mathrm{sat}}$.[1] Note that the temperatures (T_{sim}), box lengths (L_{box}), and number of molecules (N_{M}) are the exact values used in simulation while the density (ρ) is approximate (rounded) since it is calculated from L_{box}, N_{M}, and the molecular weight. Also, notice that twice as many molecules were required for *n*-octane. The reason for this is explained in Section SI.II.2.

SI.I.2 Simulation Specifications

The simulation specifications are found in Tables SI.VI-SI.VII. A LINCS-order of 8 improved results for *n*-octane, while the default value of 4 appeared to be sufficient for smaller molecules. Longer equilibration and production simulations are required for larger molecules. The output frequency is such that 1000 snapshots are sampled for each system.

Table SI.I: State points simulated for ethane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
137.0	3.21680	400	600.01
198.5	3.21680	400	600.01
174.0	3.29730	400	557.13
234.6	3.29730	400	557.13
207.0	3.38640	400	514.30
262.9	3.38640	400	514.30
236.0	3.48610	400	471.42
285.1	3.48610	400	471.42
260.0	3.59860	400	428.58
301.9	3.59860	400	428.58
360.0	6.15360	400	85.712
360.0	4.88410	400	171.43
360.0	4.26660	400	257.15
360.0	3.87650	400	342.85
360.0	3.59860	400	428.58
360.0	3.48610	400	471.42
360.0	3.38640	400	514.30
360.0	3.29730	400	557.13
360.0	3.21680	400	600.01

SI.I.3 Force Field Specifications

The intramolecular interactions used in this study are adopted from the TraPPE-UA and Potoff force fields. For the molecules studied in this work, a fixed bond-length of 0.154 nm is used for each bond between united-atom sites. Bonds are constrained using the LINCS algorithm implemented in GROMACS.

Angular bending interactions are evaluated using a harmonic potential:

$$u^{\mathrm{bend}} = \frac{k_\theta}{2}\left(\theta - \theta_0\right)^2$$

where θ is the instantaneous bond angle, θ_0 is the equilibrium bond angle, and k_θ is the harmonic force constant. For the molecules studied in this work, $\theta_0 = 114.0°$ and $k_\theta =$

Table SI.II: State points simulated for hexafluoroethane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
173.1	3.80434	400	1664.90
231.6	3.80434	400	1664.90
207.9	3.89949	400	1545.98
260.9	3.89949	400	1545.98
233.0	4.00493	400	1427.06
279.8	4.00493	400	1427.06
252.8	4.12279	400	1308.14
293.6	4.12279	400	1308.14
268.2	4.25587	400	1189.21
303.7	4.25587	400	1189.21
350.0	7.27744	400	237.84
350.0	5.77611	400	475.69
350.0	5.04589	400	713.53
350.0	4.58450	400	951.37
350.0	4.25587	400	1189.21
350.0	4.12279	400	1308.14
350.0	4.00493	400	1427.06
350.0	3.89949	400	1545.98
350.0	3.80434	400	1664.90

62500 K/rad^2 for each angle.

Dihedral torsional interactions are determined using a cosine series:

$$u^{\mathrm{tors}} = c_1[1 + \cos\phi] + c_2[1 - \cos 2\phi] + c_3[1 + \cos 3\phi]$$

where ϕ is the dihedral angle and the Fourier constants are $c_1/k_{\mathrm{B}} = 355.03$ K, $c_2/k_{\mathrm{B}} = -68.19$ K, and $c_3/k_{\mathrm{B}} = 791.32$ K.

The non-bonded force field parameters for TraPPE-UA and Potoff are provided in Table SI.VIII.

Table SI.III: State points simulated for propane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
166	3.55643	400	651.13
242	3.55643	400	651.13
210	3.64538	400	604.62
285	3.64538	400	604.62
250	3.74395	400	558.11
320	3.74395	400	558.11
285	3.85413	400	511.60
347	3.85413	400	511.60
314	3.97854	400	465.09
368	3.97854	400	465.09
444	6.80321	400	93.019
444	5.39971	400	186.04
444	4.71708	400	279.06
444	4.28575	400	372.08
444	3.97854	400	465.09
444	3.85413	400	511.60
444	3.74395	400	558.11
444	3.64538	400	604.62
444	3.55643	400	651.13

SI.I.4 GROMACS Input Files

We have provided example input files for simulating *n*-octane at 600.0 K with the TraPPE-UA force field in GROMACS (see attached .gro, .top, and .mdp files).

Table SI.IV: State points simulated for *n*-butane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
191	3.83864	400	682.53
278	3.83864	400	682.53
241	3.93465	400	633.78
327	3.93465	400	633.78
287	4.04104	400	585.03
367	4.04104	400	585.03
328	4.15997	400	536.28
399	4.15997	400	536.28
361	4.29425	400	487.52
423	4.29425	400	487.52
510	7.34306	400	97.50
510	5.82819	400	195.01
510	5.09140	400	292.51
510	4.62584	400	390.02
510	4.29425	400	487.52
510	4.15997	400	536.28
510	4.04104	400	585.03
510	3.93465	400	633.78
510	3.83864	400	682.53

Table SI.V: State points simulated for *n*-octane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
285.92	5.98449	800	708.01
387.29	5.98449	800	708.01
347.68	6.13416	800	657.44
440.25	6.13416	800	657.44
404.46	6.30003	800	606.87
483.20	6.30003	800	606.87
451.48	6.48542	800	556.30
515.25	6.48542	800	556.30
490.78	6.69481	800	505.72
539.92	6.69481	800	505.72
600.00	11.44803	800	101.14
600.00	9.08616	800	202.29
600.00	7.93753	800	303.43
600.00	7.21175	800	404.58
600.00	6.69481	800	505.72
600.00	6.48542	800	556.30
600.00	6.30003	800	606.87
600.00	6.13416	800	657.44
600.00	5.98449	800	708.01

Table SI.VI: General simulation specifications.

Integrator	Velocity Verlet
Time-step (fs)	2
Cut-off length (nm)	1.4
Thermostat	Nosé-Hoover
Time Constant (ps)	1
Constraints	LINCS
LINCS-order	8

Table SI.VII: Simulation specifications that depend on the system.

Compound	Equilibration Time (ns)	Production Time (ns)	Output Frequency (1/ps)
ethane	0.1	1	1
hexafluoroethane	0.1	1	1
propane	0.1	2	0.5
n-butane	0.2	2	0.5
n-octane	0.5	4	0.25

Table SI.VIII: Non-bonded (intermolecular) parameters for TraPPE-UA and Potoff force fields.

	TraPPE-UA			Potoff		
Interaction	ϵ (K)	σ (nm)	λ	ϵ (K)	σ (nm)	λ
$CH_3 \times CH_3$	98	0.375	12	121.25	0.3783	16
$CH_2 \times CH_2$	46	0.395	12	61	0.399	16
$CH_2 \times CH_3$	67	0.385	12	86	0.389	16
$CF_3 \times CF_3$	87	0.436	12	155.75	0.4475	36

SI.II Converting NVT simulation output to VLE properties

Section SI.II.1 provides the simulation output values for propane, *n*-butane, and *n*-octane using the TraPPE-UA and Potoff force fields. Section SI.II.2 demonstrates the ITIC data analysis for TraPPE-UA *n*-octane. Section SI.II.3 validates the TraPPE-UA and Potoff ITIC $\rho_{\mathrm{l}}^{\mathrm{sat}}$, $\rho_{\mathrm{v}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ values by comparison with those reported by Martin and Siepmann[2] and Potoff and Bernard-Brunel,[3] respectively.

SI.II.1 Simulation Results

Tables SI.IX-SI.XI contain the raw simulation U^{dep} and Z values for propane, *n*-butane, and *n*-octane, respectively. Note that each row in Tables SI.IX-SI.XI corresponds to the state point for the respective row in Tables SI.III-SI.V. We provide the results for both the TraPPE-UA and Potoff force fields. Table SI.XI also provides the 400 molecule system results for TraPPE-UA *n*-octane (where the box size is adjusted accordingly to maintain the same ρ as reported in Table SI.V for 800 molecules). Note that the standard deviations from fluctuations in a single simulation are approximately the same order of magnitude as the final digit after the decimal. A more rigorous assessment of uncertainty would be the standard deviation of the average values from several replicate simulations. We deemed this additional computational cost unnecessary for our purposes, especially since most of the uncertainty is attributed to the ITIC analysis and not the fluctuations in the simulation output.

SI.II.2 ITIC Data Analysis

Since simulations are only performed at 19 ρ–T state points, integration of Equation 5 requires some approximation. The two primary methods to integrate Equation 5 are to use numerical integration or to fit the simulation results to a model and integrate the

Table SI.IX: Raw simulation data for propane.

TraPPE-UA		Potoff	
$U^{\mathrm{dep}}\left(\frac{\mathrm{kJ}}{\mathrm{mol}}\right)$	Z	$U^{\mathrm{dep}}\left(\frac{\mathrm{kJ}}{\mathrm{mol}}\right)$	Z
-17.842	-0.253	-20.223	0.125
-16.936	3.286	-19.255	4.613
-16.126	-0.091	-18.186	0.121
-15.436	2.251	-17.482	2.942
-14.540	-0.019	-16.328	0.002
-14.058	1.451	-15.808	1.871
-13.077	-0.031	-14.598	0.010
-12.703	1.007	-14.219	1.198
-11.708	0.024	-13.016	-0.033
-11.413	0.703	-12.720	0.742
-2.419	0.751	-2.670	0.715
-4.609	0.608	-5.059	0.567
-6.692	0.531	-7.351	0.546
-8.834	0.714	-9.749	0.773
-11.067	1.347	-12.351	1.563
-12.187	1.953	-13.679	2.333
-13.236	2.886	-14.978	3.577
-14.181	4.153	-16.150	5.352
-14.920	5.947	-17.095	7.923

model. Although Razavi demonstrated that numerical integration with Simpson's rule performs well, in this study we fit simulation results to a polynomial model before performing the (exact) integration. Specifically, we fit a third order polynomial for $\frac{Z-1}{\rho}$ with respect to ρ along the isotherm. For each of the isochores, we fit Z and $\frac{U^{\mathrm{dep}}}{R_{\mathrm{g}}T}$ with respect to $\frac{1}{T}$ to a second and first order polynomial, respectively. The isotherm fit for $\frac{Z-1}{\rho}$ and the isochore fits for $\frac{U^{\mathrm{dep}}}{R_{\mathrm{g}}T}$ are used to integrate Equation 5. The isochore fits for Z are used to compute T^{sat} after recalculating $Z_{\mathrm{l}}^{\mathrm{sat}}$ with Equation 8. We performed a brief validation comparison between our ITIC analysis and that of Razavi using the same U^{dep}, Z, B_2, and B_3 values. Deviations for T^{sat} were less than 0.02 %, while deviations for $\rho_{\mathrm{v}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ were between 1-3 % (increasing percent deviations with decreasing T^{sat}).

Table SI.X: Raw simulation data for *n*-butane.

TraPPE-UA		Potoff	
$U^{\text{dep}}\left(\frac{\text{kJ}}{\text{mol}}\right)$	Z	$U^{\text{dep}}\left(\frac{\text{kJ}}{\text{mol}}\right)$	Z
-21.919	-0.214	-24.924	0.339
-20.767	3.662	-23.700	5.162
-19.744	-0.099	-22.353	0.150
-18.887	2.427	-21.441	3.225
-17.743	-0.022	-19.967	0.113
-17.105	1.632	-19.312	2.088
-15.893	0.017	-17.803	0.072
-15.430	1.086	-17.298	1.389
-14.195	0.014	-15.829	0.006
-13.822	0.780	-15.450	0.834
-2.925	0.741	-3.217	0.739
-5.557	0.558	-6.107	0.566
-8.075	0.594	-8.865	0.579
-10.660	0.746	-11.790	0.798
-13.411	1.410	-14.984	1.659
-14.809	2.062	-16.649	2.547
-16.130	3.065	-18.273	3.914
-17.305	4.569	-19.746	5.939
-18.265	6.561	-20.958	8.871

Figure SI.1 helps visualize the ITIC analysis. Panel a) demonstrates that the third order polynomial fit is adequate for interpolating the simulation results for $\frac{Z-1}{\rho}$ with respect to ρ. Sometimes a fourth order polynomial is needed to ensure adequate agreement with the REFPROP B_2 value (i.e. the intercept of $\frac{Z-1}{\rho}$ with respect to ρ). Panel b) shows that a first order polynomial is sufficient for fitting $\frac{U^{\text{dep}}}{R_{\text{g}}T}$ with respect to $\frac{1}{T}$. Panel c) demonstrates that a second order polynomial accurately fits Z with respect to $\frac{1}{T}$. However, to ensure the correct concavity, we used a first order polynomial whenever the second order fit was concave up.

The results presented in Figure SI.1 are for the specific example of *n*-octane using the TraPPE-UA model. This molecule was selected for at least two reasons. First, it is the

Table SI.XI: Raw simulation data for *n*-octane.

TraPPE-UA, 800 molecules		Potoff, 800 molecules		TraPPE-UA, 400 molecules	
$U^{\text{dep}}\left(\frac{\text{kJ}}{\text{mol}}\right)$	Z	$U^{\text{dep}}\left(\frac{\text{kJ}}{\text{mol}}\right)$	Z	$U^{\text{dep}}\left(\frac{\text{kJ}}{\text{mol}}\right)$	Z
-22.707	0.032	-25.444	0.071	-22.701	-0.222
-22.286	0.546	-24.980	0.620	-22.252	0.201
-25.486	-0.023	-28.678	0.033	-25.497	-0.255
-24.913	0.911	-28.053	1.080	-24.900	0.643
-28.551	-0.041	-32.268	0.088	-28.582	-0.414
-27.736	1.470	-31.397	1.844	-27.751	1.134
-31.908	-0.169	-36.273	0.104	-32.084	-0.941
-30.750	2.333	-35.030	3.080	-30.791	1.952
-35.606	-0.369	-40.679	0.210	-36.092	-1.990
-33.983	3.669	-38.902	5.206	-34.115	2.953
-5.177	0.663	-5.850	0.670	-5.088	0.510
-9.501	0.421	-10.597	0.443	-9.391	0.322
-13.353	0.341	-14.856	0.320	-13.271	0.129
-17.353	0.422	-19.298	0.429	-17.280	0.362
-21.823	1.098	-24.471	1.306	-21.814	0.891
-24.246	1.783	-27.350	2.219	-24.217	1.475
-26.675	2.916	-30.273	3.600	-26.654	2.849
-29.022	4.591	-33.183	5.914	-28.997	4.357
-31.141	6.916	-35.826	9.407	-31.090	6.733

only molecule studied that has angular, torsional, and non-bonded intramolecular (i.e. 1-5, 1-6, 1-7, and 1-8) interactions. U^{dep} should not include these non-bonded intramolecular interactions. Unfortunately, GROMACS does not distinguish between intramolecular and intermolecular non-bonded energies in the output files. For this reason, U^{dep} was determined by performing "rerun" simulations where intramolecular non-bonded interactions were excluded but the configurations were those sampled using the full force field potential. Second, we observed that for larger molecules, such as *n*-octane, 400 molecule simulations did not provide reliable values for Z (see Table SI.XIV). This resulted in extremely poor VLE estimates compared to those reported by Martin and Siepmann[2] (see Table SI.XV). Figure SI.1 Panel a) includes the $\frac{Z-1}{\rho}$ values for the 400 molecule simulations

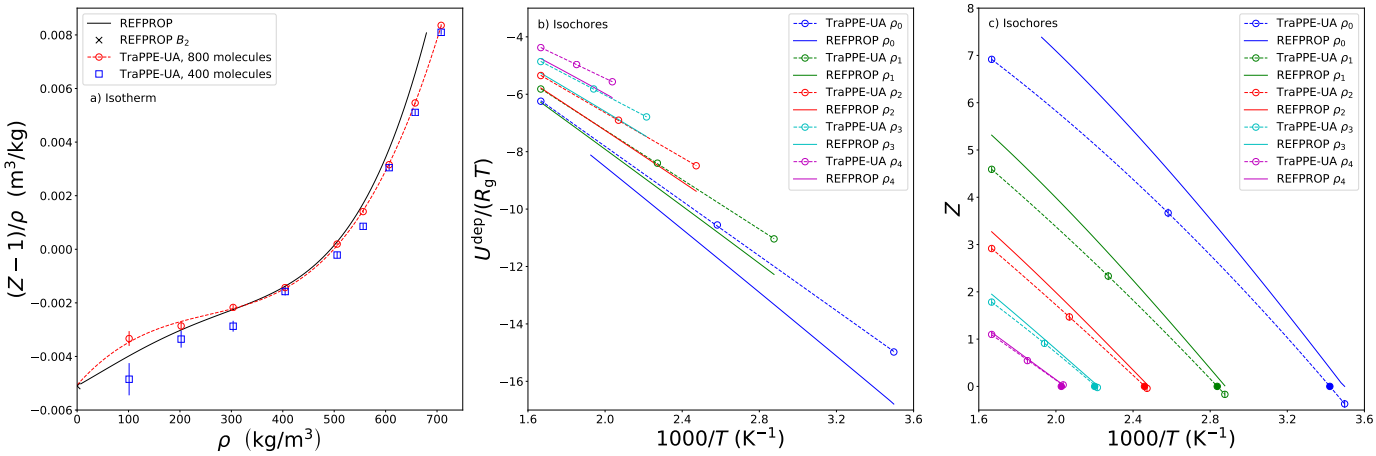

Figure SI.1: Simulation results at ITIC state points for *n*-octane using the TraPPE-UA model. Raw simulation values are depicted as open symbols while dashed lines are the polynomial fits to the simulation output. REFPROP values are included as solid lines for comparison. Panel a) plots the isotherm results for $\frac{Z-1}{\rho}$ with respect to ρ for the 400 and 800 molecule simulations. Panels b) and c) plot, respectively, the isochore results for $\frac{U^{\text{dep}}}{R_{\text{g}}T}$ and Z with respect to $\frac{1000}{T}$ for the 800 molecule systems. $\rho_0 = 708.00\frac{\text{kg}}{\text{m}^3}, \rho_1 = 657.43\frac{\text{kg}}{\text{m}^3}, \rho_2 = 606.86\frac{\text{kg}}{\text{m}^3}, \rho_3 = 556.29\frac{\text{kg}}{\text{m}^3}, \rho_4 = 505.71\frac{\text{kg}}{\text{m}^3}$. Panel a) includes the B_2 REFPROP value for the isotherm as the intercept. Filled symbols in Panel c) correspond to the interpolated T^{sat} values. Error bars represent two times the standard deviation of the fluctuations from a single simulation.

to demonstrate this system-size dependence.

SI.II.3 Validation

Figure SI.2 compares the ITIC results with the literature VLE values reported for the TraPPE-UA[2] and Potoff force fields.[3] The ITIC $\rho_{\text{l}}^{\text{sat}}$ and $P_{\text{v}}^{\text{sat}}$ values agree within 1% and 5%, respectively, with those obtained using more traditional VLE methods, such as Gibbs Ensemble Monte Carlo and Grand Canonical Monte Carlo. Panels b) and d) also demonstrate that the ITIC $\rho_{\text{l}}^{\text{sat}}$ and $P_{\text{v}}^{\text{sat}}$ values have a relatively small dependence on B_2 and B_3. Specifically, the ITIC results are similar when using REFPROP B_2 and B_3 values or the more rigorous Mayer-sampling Monte Carlo B_2 and B_3 values reported by Schultz and Kofke for the TraPPE-UA model[4]

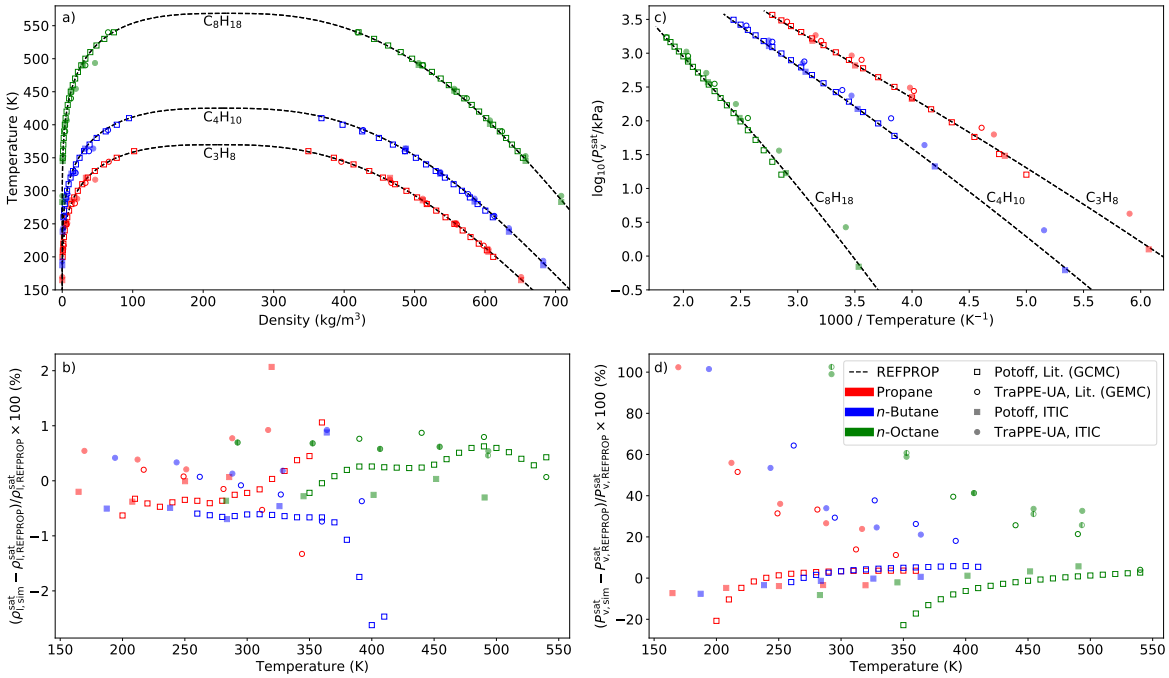

Figure SI.2: ITIC results are in good agreement with literature values for Potoff and TraPPE-UA models. Using REFPROP B_2 and B_3 introduces a reasonably small error into the ITIC P_v^{sat} results. Panels a)-b) plot ρ_l^{sat} and ρ_v^{sat} while Panels c)-d) plot P_v^{sat}. Panels b) and d) are percent deviation plots of simulation results relative to REFPROP values. Square and circle symbols correspond to the Potoff and TraPPE-UA force fields, respectively. Empty symbols are literature values (Potoff and Bernard-Brunel[3] and Martin and Siepmann[2]) while filled symbols are from this study (ITIC). Red, blue, and green symbols correspond to propane, *n*-butane, and *n*-octane, respectively. Panels b) and d) include half-filled green circles that are ITIC results obtained from the same simulation data as the filled green circles (TraPPE-UA *n*-octane) but using Mayer-sampling Monte Carlo B_2 and B_3 values[4] instead of the REFPROP B_2 and B_3 values.

Tables SI.XII-SI.XIV provide the ITIC T^{sat}, ρ_l^{sat}, ρ_v^{sat}, and P_v^{sat} values for propane, *n*-butane, and *n*-octane, respectively, for both the TraPPE-UA and Potoff force fields. Table SI.XV provides the TraPPE-UA *n*-octane ITIC results when using the Schultz B_2 and B_3 values[4] instead of REFPROP and for the 400 molecule system simulations. Notice that the ρ_l^{sat} values in Tables SI.XII-SI.XV for both TraPPE-UA and Potoff are exactly the same as the simulated isochore densities (see Tables SI.III-SI.V). This is expected because ITIC

solves for T^{sat} for a fixed $\rho_{\mathrm{l}}^{\mathrm{sat}}$. Also, note that no attempt was made to quantify the uncertainty in the T^{sat}, $\rho_{\mathrm{v}}^{\mathrm{sat}}$, and $P_{\mathrm{v}}^{\mathrm{sat}}$ values. Quantifying the ITIC uncertainties is not as straightforward as other VLE methods (i.e. GEMC, GCMC). Our preliminary recommendation is to bootstrap the uncertainties by performing around five replicate simulations at each state point and randomly selecting which replicates are included in the ITIC analysis.

Table SI.XII: ITIC results for propane.

TraPPE-UA				Potoff			
T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)	T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)
169.4	651.13	0.042	0.133	164.7	651.13	0.013	0.041
212.1	604.62	0.628	1.620	207.8	604.62	0.304	0.789
251.1	558.11	3.092	7.193	250.1	558.11	2.108	4.763
288.0	511.60	9.235	21.208	285.6	511.60	6.584	14.147
316.8	465.09	18.451	46.824	319.6	465.09	15.303	33.591

Table SI.XIII: ITIC results for *n*-butane.

TraPPE-UA				Potoff			
T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)	T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)
194.0	682.53	0.024	0.087	187.3	682.53	0.006	0.023
243.4	633.78	0.438	1.291	238.1	633.78	0.212	0.629
288.1	585.03	2.359	6.243	283.8	585.03	1.498	3.900
328.6	536.28	7.096	18.383	325.9	536.28	5.313	13.101
364.0	487.52	15.397	44.023	363.9	487.52	12.757	32.431

Table SI.XIV: ITIC results for *n*-octane using 800 molecule simulations with REFPROP B_2 and B_3 values.

TraPPE-UA				Potoff			
T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)	T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)
292.3	708.00	0.025	0.120	283.0	708.00	0.007	0.033
352.5	657.43	0.345	1.385	345.2	657.43	0.167	0.675
406.6	606.86	1.673	6.210	401.4	606.86	1.099	4.000
454.3	556.29	4.766	17.676	451.6	556.29	3.733	13.223
493.1	505.71	9.595	38.943	490.5	505.71	8.002	29.747

Table SI.XV: TraPPE-UA ITIC results for *n*-octane using Schultz B_2 and B_3 values and 400 molecule simulations. Compare with TraPPE-UA values reported in Table SI.XIV.

TraPPE-UA, Schultz B_2 and B_3				TraPPE-UA, 400 molecules			
T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)	T^{sat} (K)	$\rho_{\mathrm{l}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$\rho_{\mathrm{v}}^{\mathrm{sat}}\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$	$P_{\mathrm{v}}^{\mathrm{sat}}$ (bar)
292.3	708.00	0.027	0.128	317.0	708.00	0.062	0.269
352.5	657.43	0.367	1.464	371.8	657.43	0.477	1.825
406.6	606.86	1.771	6.459	423.6	606.86	1.858	6.584
454.4	556.29	5.005	17.862	469.1	556.29	4.612	15.939
493.2	505.71	9.946	37.074	517.8	505.71	9.992	35.316

SI.III Alternative PCFR-optimal

Although we recommend minimizing the RMS deviations from REFPROP $\log_{10}\left(P_{\mathrm{v}}^{\mathrm{sat}}\right)$ values to determine the PCFR-optimal ϵ for a given λ, minimizing the RMS of U^{dep} is also a reliable method. However, since $P_{\mathrm{v}}^{\mathrm{sat}}$ experimental data are more readily available than U^{dep}, an optimization based on U^{dep} is not useful for most compounds.

Figure SI.3 presents the PCFR-optimal points based on minimizing the RMS for U^{dep}. The trends in Panel a) for CH_3 are similar to those observed in Figure 8 for $\log_{10}\left(P_{\mathrm{v}}^{\mathrm{sat}}\right)$. The CH_2 results in Panel b) suggest that a better $\epsilon_{\mathrm{ref,CH_2}}$ value for $\lambda_{\mathrm{ref,CH_2}} = 16$ would be around 61 K, compared to the value of 54 K obtained from $\log_{10}\left(P_{\mathrm{v}}^{\mathrm{sat}}\right)$. Fortunately, MBAR does not require an extremely accurate estimate of ϵ_{ref} to yield reliable results.

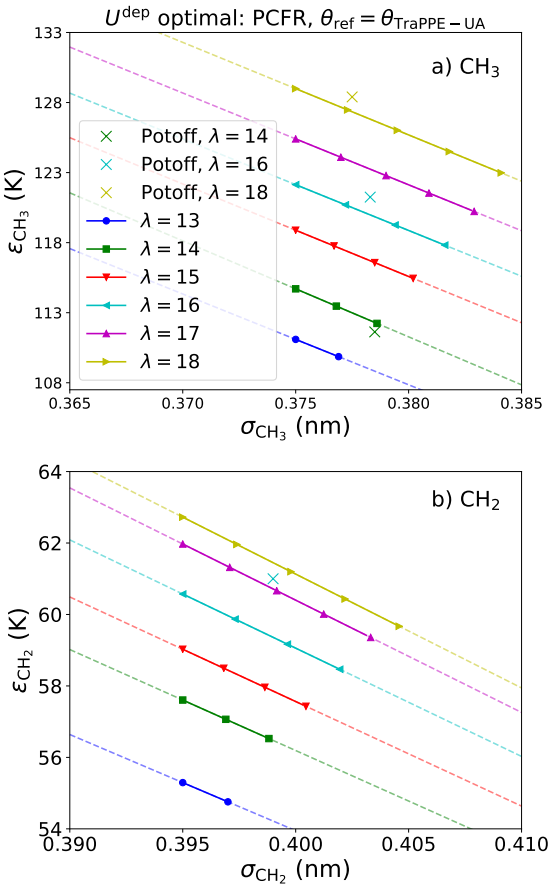

Figure SI.3: Optimized ϵ and σ sets for different values of λ based on the PCFR RMS of U^{dep}. Panels a) and b) correspond to CH_3 and CH_2 UA sites, respectively. The transparent dashed lines are the entire range of optimal values while the opaque solid lines are constrained by σ and r_{min}. The points along the lines represent ϵ and σ sets to sample for a given λ.

SI.IV Basis Functions

SI.IV.1 Theory

In this study, we utilize Mie λ-6 basis functions. Basis functions are a tool for reducing the computational cost of the energy and pressure (virial) "rerun" calculations that are required for MBAR. Basis functions for non-bonded potentials can be developed when the interaction energies and forces are linear functions of the non-bonded parameters (or a set of non-bonded parameters). For example, basis functions can be derived by using an alternative expression for the Mie λ-6 potential:

$$u^{\mathrm{vdw}}(C_6, C_\lambda; r) = C_\lambda r^{-\lambda} - C_6 r^{-6}$$

$$\frac{\partial u^{\mathrm{vdw}}}{\partial r}(C_6, C_\lambda; r) = -\lambda C_\lambda r^{-\lambda-1} + 6C_6 r^{-7}$$

where C_6 and C_λ are proportional to $\epsilon\sigma^6$ and $\epsilon\sigma^\lambda$, respectively. The internal energy (U^{vdw}) and internal virial (W^{vdw})[5] from the non-bonded interactions can then be expressed as:

$$U^{\mathrm{vdw}}(C_6, C_\lambda) = C_\lambda \sum_{i=1}^{N_\mathrm{S}-1} \sum_{j>i}^{N_\mathrm{S}} r_{ij}^{-\lambda} - C_6 \sum_{i=1}^{N_\mathrm{S}-1} \sum_{j>i}^{N_\mathrm{S}} r_{ij}^{-6}$$

$$-3W^{\mathrm{vdw}}(C_6, C_\lambda) = -\lambda C_\lambda \sum_{i=1}^{N_\mathrm{S}-1} \sum_{j>i}^{N_\mathrm{S}} r_{ij}^{-\lambda} + 6C_6 \sum_{i=1}^{N_\mathrm{S}-1} \sum_{j>i}^{N_\mathrm{S}} r_{ij}^{-6}$$

The double summations over the interacting site types can be computationally demanding as the number of interaction sites (N_S) increases, i.e. for larger compounds and/or simulations with more molecules. However, notice that the double summations depend only on the distance (r_{ij}) between sites i and j and not on C_6 and C_λ (i.e. the Mie λ-6 parameters). When the C_6 and C_λ parameters (or ϵ and σ) are varied, the double summations do not need to be recomputed. Therefore, if the values of the double summa-

tions for $r_{ij}^{-\lambda}$ and r_{ij}^{-6} (i.e. basis functions) are stored at each snapshot, $U^{\mathrm{vdw}}(C_6, C_\lambda)$ and $W^{\mathrm{vdw}}(C_6, C_\lambda)$ (or alternatively $U^{\mathrm{dep}}(\theta)$ and $Z(\theta)$, respectively) can be computed without any "rerun" calculations.

In addition, the memory storage requirement for basis functions is much less compared to storing the configurations $(\boldsymbol{x}_n)$ for N snapshots (which must be full-precision for reliable "rerun" results). In fact, the memory required for the basis function approach is constant with respect to N_{S}. Therefore, the memory saved becomes substantial for larger compounds and/or simulations with more molecules. For example, 1001 snapshots of 800 *n*-octane UA molecules requires approximately 0.4 MB and 150 MB of file storage with and without basis functions, respectively.

SI.IV.2 Implementation

Although the equations in Section SI.IV.1 are straightforward in theory, a significant amount of post-processing is necessary to develop basis functions. This is because GROMACS (and any other open-source package that we are aware of) does not provide basis functions as output. Future collaboration with software developers to include an option for returning basis functions will facilitate use of this approach. This section presents the brute-force algorithm we used for converting the GROMACS output into basis functions.

GROMACS provides the energies from bond stretching, angles bending, dihedral torsions and non-bonded interactions as well as the virial tensor. However, GROMACS does not distinguish between the contributions arising from different site-types (i.e. the non-bonded CH_3, CH_2, or cross-interactions) which is necessary to develop basis functions for each site-type. Also, the non-bonded energies provided by GROMACS include both intermolecular and intramolecular (i.e. 1-5, 1-6, etc.) interactions. The ITIC approach requires U^{dep}, which includes only the intermolecular contributions,[1] while MBAR utilizes the total internal energy (U) for reweighting (Equations 10-11). Therefore, molecules that

include intramolecular non-bonded interactions (such as *n*-octane) require a basis function for both the intermolecular and intramolecular non-bonded energies. (Note that we do not need basis functions for the angular and torsional contributions to the internal energy and virial because the intramolecular potential is the same for θ and θ_{ref} in this study.) Accounting for the constraint forces in the virial is another complication (this is one disadvantage of using MD instead of MC).

Rerun Calculations

Table SI.XVI summarizes the required "reruns" for developing Mie λ-6 basis functions for *n*-octane. This set of 19 "rerun" calculations is performed for every ITIC state-point (ρ-T). Although this can be somewhat time-consuming, these "reruns" only need to be performed once to generate the basis functions. Therefore, this significant upfront cost is negligible compared to the speed-up achieved with the basis functions.

To simplify the notation, non-bonded interaction types are labeled as CH_3, CH_2, and CX for CH_3–CH_3, CH_2–CH_2, and CH_3–CH_2 (cross-interactions). The respective $C_{6,i}^{\text{low}}$, $C_{\lambda,i}^{\text{low}}$, $C_{6,i}^{\text{high}}$, and $C_{\lambda,i}^{\text{high}}$ values used in this study for the Mie 16-6 potential are provided in Table SI.XVII. The exact value for these variables is not important as long as they are in a reasonable range. We use "low" and "high" to refer to C_6 and C_λ values that are determined from the lower and higher bounds of the investigated ϵ and σ parameter space (i.e. $C_{\lambda,\text{CH}_2}^{\text{low}} \propto 50 \times 0.385^{16}$ K nm^{16} for $\lambda_{\text{CH}_2} = 16$).

Rather than using two different $C_{6,i}$ and $C_{\lambda,i}$ sets (i.e. $C_{6,i}^{\text{low}}$, $C_{\lambda,i}^{\text{low}}$, and $C_{6,i}^{\text{high}}$, $C_{\lambda,i}^{\text{high}}$) for CH_3, CH_2, and CX interactions, it is also possible to use a single $C_{6,i}$ and $C_{\lambda,i}$ value with the other equal to zero. Although this approach might appear more straightforward and conceptually simple, this requires the same number of "reruns" and typically has worse precision as the magnitude of the energies and forces becomes very large. This is because setting C_6 or C_λ to zero means that only repulsive or attractive interactions are included.

Table SI.XVI: Required "rerun" calculations to develop Mie λ-6 basis functions for *n*-octane using GROMACS.

Index	Constraints	Exclusions	$C_{6,\mathrm{CH_3}}$	$C_{\lambda,\mathrm{CH_3}}$	$C_{6,\mathrm{CH_2}}$	$C_{\lambda,\mathrm{CH_2}}$	$C_{6,\mathrm{CX}}$	$C_{\lambda,\mathrm{CX}}$
0	Yes	3	0	0	0	0	0	0
1	Yes	3	$C_{6,\mathrm{CH_3}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}}$	0	0	0	0
2	Yes	3	$C_{6,\mathrm{CH_3}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}}$	0	0	0	0
3	No	3	$C_{6,\mathrm{CH_3}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}}$	0	0	0	0
4	No	3	$C_{6,\mathrm{CH_3}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}}$	0	0	0	0
5	Yes	8	$C_{6,\mathrm{CH_3}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}}$	0	0	0	0
6	Yes	8	$C_{6,\mathrm{CH_3}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}}$	0	0	0	0
7	Yes	3	0	0	$C_{6,\mathrm{CH_2}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{low}}$	0	0
8	Yes	3	0	0	$C_{6,\mathrm{CH_2}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{high}}$	0	0
9	No	3	0	0	$C_{6,\mathrm{CH_2}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{low}}$	0	0
10	No	3	0	0	$C_{6,\mathrm{CH_2}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{high}}$	0	0
11	Yes	8	0	0	$C_{6,\mathrm{CH_2}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{low}}$	0	0
12	Yes	8	0	0	$C_{6,\mathrm{CH_2}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{high}}$	0	0
13	Yes	3	0	0	0	0	$C_{6,\mathrm{CX}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{low}}$
14	Yes	3	0	0	0	0	$C_{6,\mathrm{CX}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{high}}$
15	No	3	0	0	0	0	$C_{6,\mathrm{CX}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{low}}$
16	No	3	0	0	0	0	$C_{6,\mathrm{CX}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{high}}$
17	Yes	8	0	0	0	0	$C_{6,\mathrm{CX}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{low}}$
18	Yes	8	0	0	0	0	$C_{6,\mathrm{CX}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{high}}$

Table SI.XVII: Values used in this study for C_6 and C_λ in "rerun" calculations with $\lambda = 16$ for CH_3, CH_2, and CX interactions. Units are kJ/mol.

$C_{6,\mathrm{CH_3}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}}$	$C_{6,\mathrm{CH_2}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{low}}$	$C_{6,\mathrm{CX}}^{\mathrm{low}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{low}}$
0.006119503	2.568294E-07	0.003901841	2.791752E-07	0.004896883	2.692975E-07
$C_{6,\mathrm{CH_3}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}}$	$C_{6,\mathrm{CH_2}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CH_2}}^{\mathrm{high}}$	$C_{6,\mathrm{CX}}^{\mathrm{high}}$	$C_{\lambda,\mathrm{CX}}^{\mathrm{high}}$
0.009988713	7.146886E-07	0.007402244	8.788477E-07	0.008615327	7.966045E-07

Data Analysis

We now demonstrate how to convert the "rerun" GROMACS results into basis functions. The following equations are written for analyzing a single snapshot. Therefore, this

procedure must be repeated for all N snapshots. For brevity, we only demonstrate this approach in detail for the CH_3 interactions. The process for developing the CH_2 and CX basis functions is discussed briefly afterwards.

A basis function $(\mathbf{\Psi})$ for property value (Y) and interaction site-type (X) is obtained by solving a linear system of equations:

$$\mathbf{Y} = \mathbf{C}_X \mathbf{\Psi}_X^Y$$

where $\mathbf{C}_X$ is a matrix of C_6 and C_λ parameters for X and $\mathbf{Y}$ is an array of the corresponding property values. By solving this equation for $\mathbf{\Psi}_X^Y$ we obtain:

$$\mathbf{\Psi}_X^Y = \left(\mathbf{C}_X^T \mathbf{C}_X\right)^{-1} \left(\mathbf{C}_X^T \mathbf{Y}\right)$$

For example, the basis function for the non-bonded intermolecular energy contributions from CH_3 interactions is obtained from:

$$\begin{bmatrix} \Psi_{6,\mathrm{CH_3}}^{U^{\mathrm{vdw,inter}}} \\ \Psi_{\lambda,\mathrm{CH_3}}^{U^{\mathrm{vdw,inter}}} \end{bmatrix} = \left(\begin{bmatrix} C_{6,\mathrm{CH_3}}^{\mathrm{low}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}} \\ C_{6,\mathrm{CH_3}}^{\mathrm{high}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}} \end{bmatrix}^T \begin{bmatrix} C_{6,\mathrm{CH_3}}^{\mathrm{low}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}} \\ C_{6,\mathrm{CH_3}}^{\mathrm{high}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}} \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} C_{6,\mathrm{CH_3}}^{\mathrm{low}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}} \\ C_{6,\mathrm{CH_3}}^{\mathrm{high}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}} \end{bmatrix}^T \begin{bmatrix} U_{\mathrm{rr=5}}^{\mathrm{vdw}} \\ U_{\mathrm{rr=6}}^{\mathrm{vdw}} \end{bmatrix} \right)$$

where the subscripts "rr=5" and "rr=6" refer to the "rerun" index in Table SI.XVI. Similarly, the basis function for the non-bonded intramolecular energy contributions from CH_3 interactions is obtained from:

$$\begin{bmatrix} \Psi_{6,\mathrm{CH_3}}^{U^{\mathrm{vdw,intra}}} \\ \Psi_{\lambda,\mathrm{CH_3}}^{U^{\mathrm{vdw,intra}}} \end{bmatrix} = \left(\begin{bmatrix} C_{6,\mathrm{CH_3}}^{\mathrm{low}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}} \\ C_{6,\mathrm{CH_3}}^{\mathrm{high}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}} \end{bmatrix}^T \begin{bmatrix} C_{6,\mathrm{CH_3}}^{\mathrm{low}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}} \\ C_{6,\mathrm{CH_3}}^{\mathrm{high}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}} \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} C_{6,\mathrm{CH_3}}^{\mathrm{low}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{low}} \\ C_{6,\mathrm{CH_3}}^{\mathrm{high}} & C_{\lambda,\mathrm{CH_3}}^{\mathrm{high}} \end{bmatrix}^T \begin{bmatrix} U_{\mathrm{rr=1}}^{\mathrm{vdw}} - U_{\mathrm{rr=5}}^{\mathrm{vdw}} \\ U_{\mathrm{rr=2}}^{\mathrm{vdw}} - U_{\mathrm{rr=6}}^{\mathrm{vdw}} \end{bmatrix} \right)$$

The basis function for the non-bonded virial contributions from CH_3 interactions is ob-

tained from:

$$\begin{bmatrix} \Psi^{W^{\mathrm{vdw}}}_{6,\mathrm{CH_3}} \\ \Psi^{W^{\mathrm{vdw}}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} = \left(\begin{bmatrix} C^{\mathrm{low}}_{6,\mathrm{CH_3}} & C^{\mathrm{low}}_{\lambda,\mathrm{CH_3}} \\ C^{\mathrm{high}}_{6,\mathrm{CH_3}} & C^{\mathrm{high}}_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} C^{\mathrm{low}}_{6,\mathrm{CH_3}} & C^{\mathrm{low}}_{\lambda,\mathrm{CH_3}} \\ C^{\mathrm{high}}_{6,\mathrm{CH_3}} & C^{\mathrm{high}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} C^{\mathrm{low}}_{6,\mathrm{CH_3}} & C^{\mathrm{low}}_{\lambda,\mathrm{CH_3}} \\ C^{\mathrm{high}}_{6,\mathrm{CH_3}} & C^{\mathrm{high}}_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} W_{\mathrm{rr}=3} \\ W_{\mathrm{rr}=4} \end{bmatrix} \right)$$

The basis functions for the LINCS virial contributions from CH_3 non-bonded interactions is obtained from:

$$\begin{bmatrix} \Psi^{W^{\mathrm{vdw,LINCS}}}_{6,\mathrm{CH_3}} \\ \Psi^{W^{\mathrm{vdw,LINCS}}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} = \left(\begin{bmatrix} C^{\mathrm{low}}_{6,\mathrm{CH_3}} & C^{\mathrm{low}}_{\lambda,\mathrm{CH_3}} \\ C^{\mathrm{high}}_{6,\mathrm{CH_3}} & C^{\mathrm{high}}_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} C^{\mathrm{low}}_{6,\mathrm{CH_3}} & C^{\mathrm{low}}_{\lambda,\mathrm{CH_3}} \\ C^{\mathrm{high}}_{6,\mathrm{CH_3}} & C^{\mathrm{high}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} C^{\mathrm{low}}_{6,\mathrm{CH_3}} & C^{\mathrm{low}}_{\lambda,\mathrm{CH_3}} \\ C^{\mathrm{high}}_{6,\mathrm{CH_3}} & C^{\mathrm{high}}_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} W_{\mathrm{rr}=1} - W_{\mathrm{rr}=3} - W_{\mathrm{rr}=0} \\ W_{\mathrm{rr}=2} - W_{\mathrm{rr}=4} - W_{\mathrm{rr}=0} \end{bmatrix} \right)$$

Similar expressions are obtained for CH_2 and CX by adding 6 and 12, respectively, to all of the "rerun" indices greater than 0 in the equations above.

After solving these equations, the basis functions can be used to calculate the internal energy and virial for any set of C_6 and C_λ parameters. Specifically, the total non-bonded intermolecular energy (which is assumed to be equal to U^{dep}) for a set of non-bonded parameters is obtained by combining the contributions from each interaction site-type:

$$U^{\mathrm{dep}}(\theta) = U^{\mathrm{vdw,inter}}(\theta) = \begin{bmatrix} C_{6,\mathrm{CH_3}} \\ C_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} \Psi^{U^{\mathrm{vdw,inter}}}_{6,\mathrm{CH_3}} \\ \Psi^{U^{\mathrm{vdw,inter}}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CH_2}} \\ C_{\lambda,\mathrm{CH_2}} \end{bmatrix}^T \begin{bmatrix} \Psi^{U^{\mathrm{vdw,inter}}}_{6,\mathrm{CH_2}} \\ \Psi^{U^{\mathrm{vdw,inter}}}_{\lambda,\mathrm{CH_2}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CX}} \\ C_{\lambda,\mathrm{CX}} \end{bmatrix}^T \begin{bmatrix} \Psi^{U^{\mathrm{vdw,inter}}}_{6,\mathrm{CX}} \\ \Psi^{U^{\mathrm{vdw,inter}}}_{\lambda,\mathrm{CX}} \end{bmatrix}$$

Similarly, the total non-bonded intramolecular energy is:

$$U^{\mathrm{vdw,intra}}(\theta) = \begin{bmatrix} C_{6,\mathrm{CH_3}} \\ C_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} \Psi^{U^{\mathrm{vdw,intra}}}_{6,\mathrm{CH_3}} \\ \Psi^{U^{\mathrm{vdw,intra}}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CH_2}} \\ C_{\lambda,\mathrm{CH_2}} \end{bmatrix}^T \begin{bmatrix} \Psi^{U^{\mathrm{vdw,intra}}}_{6,\mathrm{CH_2}} \\ \Psi^{U^{\mathrm{vdw,intra}}}_{\lambda,\mathrm{CH_2}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CX}} \\ C_{\lambda,\mathrm{CX}} \end{bmatrix}^T \begin{bmatrix} \Psi^{U^{\mathrm{vdw,intra}}}_{6,\mathrm{CX}} \\ \Psi^{U^{\mathrm{vdw,intra}}}_{\lambda,\mathrm{CX}} \end{bmatrix}$$

The total internal energy is obtained by adding the angular and torsional contributions (which are constant when the intramolecular potential for θ does not change) to the non-

bonded contributions (the bond energy, $U^{\mathrm{bond}} = 0$ in this study):

$$U(\theta) = U^{\mathrm{bend}} + U^{\mathrm{tors}} + U^{\mathrm{vdw,inter}}(\theta) + U^{\mathrm{vdw,intra}}(\theta)$$

The total non-constrained non-bonded contribution to the virial is:

$$W^{\mathrm{vdw}}(\theta) = \begin{bmatrix} C_{6,\mathrm{CH_3}} \\ C_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} \Psi^{W^{\mathrm{vdw}}}_{6,\mathrm{CH_3}} \\ \Psi^{W^{\mathrm{vdw}}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CH_2}} \\ C_{\lambda,\mathrm{CH_2}} \end{bmatrix}^T \begin{bmatrix} \Psi^{W^{\mathrm{vdw}}}_{6,\mathrm{CH_2}} \\ \Psi^{W^{\mathrm{vdw}}}_{\lambda,\mathrm{CH_2}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CX}} \\ C_{\lambda,\mathrm{CX}} \end{bmatrix}^T \begin{bmatrix} \Psi^{W^{\mathrm{vdw}}}_{6,\mathrm{CX}} \\ \Psi^{W^{\mathrm{vdw}}}_{\lambda,\mathrm{CX}} \end{bmatrix}$$

The total constrained (LINCS) non-bonded contribution to the virial is:

$$W^{\mathrm{vdw,LINCS}}(\theta) = \begin{bmatrix} C_{6,\mathrm{CH_3}} \\ C_{\lambda,\mathrm{CH_3}} \end{bmatrix}^T \begin{bmatrix} \Psi^{W^{\mathrm{vdw,LINCS}}}_{6,\mathrm{CH_3}} \\ \Psi^{W^{\mathrm{vdw,LINCS}}}_{\lambda,\mathrm{CH_3}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CH_2}} \\ C_{\lambda,\mathrm{CH_2}} \end{bmatrix}^T \begin{bmatrix} \Psi^{W^{\mathrm{vdw,LINCS}}}_{6,\mathrm{CH_2}} \\ \Psi^{W^{\mathrm{vdw,LINCS}}}_{\lambda,\mathrm{CH_2}} \end{bmatrix} + \begin{bmatrix} C_{6,\mathrm{CX}} \\ C_{\lambda,\mathrm{CX}} \end{bmatrix}^T \begin{bmatrix} \Psi^{W^{\mathrm{vdw,LINCS}}}_{6,\mathrm{CX}} \\ \Psi^{W^{\mathrm{vdw,LINCS}}}_{\lambda,\mathrm{CX}} \end{bmatrix}$$

The total virial is:

$$W(\theta) = W^{\mathrm{vdw}}(\theta) + W^{\mathrm{vdw,LINCS}}(\theta) + W_{rr=0}$$

where $W_{rr=0}$ accounts for the constraint forces (LINCS) not related to the non-bonded interactions. $W_{rr=0}$ does not depend on θ because we assume the intramolecular potential is constant.

Finally, the virial is converted into pressure using:[6]

$$P(\theta) = \frac{2}{V}\left(\frac{KE}{3} - W(\theta)\right)$$

where the volume (V) and kinetic energy (KE) are constant with respect to the force field. Note that KE is either the kinetic energy from the corresponding snapshot or the average KE of the NVT ensemble. Although using the average KE can reduce the noise in P, since the fluctuations in W are much greater than KE the difference between the

two methods is negligible. Also, note that the long-range tail corrections to energies and pressures can be grouped into the non-bonded basis function or added analytically afterwards.

References

(1) Razavi, S. M. Optimization of a Transferable Shifted Force Field for Interfaces and Inhomogenous Fluids using Thermodynamic Integration. M.Sc. thesis, The University of Akron, 2016.

(2) Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.

(3) Potoff, J. J.; Bernard-Brunel, D. A. Mie Potentials for Phase Equilibria Calculations: Applications to Alkanes and Perfluoroalkanes. *J. Phys. Chem. B* **2009**, *113*, 14725–14731.

(4) Schultz, A. J.; Kofke, D. A. Virial coefficients of model alkanes. *J. Chem. Phys.* **2010**, *133*, 104101.

(5) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press ; Oxford University Press: Oxford England New York, 1987; pp xix, 385 p.

(6) Abraham, M.; van der Spoel, D.; Lindahl, E.; B.Hess,; the GROMACS development team, GROMACS User Manual version 2016.3, www.gromacs.org (2017).