Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2023

Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

[a]*Institute of Theoretical and Computational Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China*
[b]*Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, 19 Xin-Jie-Kou-Wai Street, Beijing 100875, China*
*Corresponding author email address: Lianghui Gao (lhgao@bnu.edu.cn); Weihai Fang (fangwh@bnu.edu.cn)

Supplementary Material: A Top-down and Bottom-up Combined Strategy for Parameterization of Coarse-grained Force Fields for Phospholipids

Mingwei Wan,[ab] Junjie Song,[b] Ying Yang,[b] Lianghui Gao,*[b] and Weihai Fang*[ab]

SI.1. OPT1.35 FF and its predictions.

Table SI-1. Bond length and non-bonded parameters.

Pairs	L_b (nm)	ε(kJ/mol)	R_0 (nm)	α	β
[a]C_3-C_3	0.410	3.091	0.534	11.442	11.907
[a]C_4-C_4	0.450	4.357	0.578	12.798	14.254
C_5-C_5	0.490	4.820	0.613	14.965	12.754
[a]C_3-C_4	0.430	3.671	0.556	12.101	13.028
C_3-C_5	0.450	4.582	0.596	13.839	13.483
C_4-C_5	0.470	3.859	0.574	13.085	12.323
[b]W_d-W_d	0.5054	3.015	0.4969	6.249	9.253
C_3-W_d		2.470	0.486	8.456	10.496
[b]C_4-W_d		2.935	0.492	8.943	11.484
C_5-W_d		3.017	0.503	9.670	10.863
COC-COC	0.364	4.033	0.512	11.392	15.557
C_3-COC	0.387	3.260	0.524	11.417	13.610
C_4-COC	0.407	3.844	0.546	12.075	14.891
C_5-COC	0.427	4.139	0.558	13.057	14.086
W_d-COC		4.137	0.492	8.437	11.998
NC_3-NC_3	0.450	4.632	0.574	13.678	14.314
C_3-NC_3	0.430	3.671	0.540	12.510	13.055
C_4-NC_3	0.450	4.501	0.567	13.231	14.284
C_5-NC_3	0.470	4.696	0.581	14.307	13.512
W_d-NC_3		4.444	0.478	9.245	11.509
COC-NC_3		4.508	0.528	12.483	14.923
COOC-COOC	0.392	5.108	0.513	7.638	17.538
C_3-COOC	0.401	3.334	0.559	9.348	14.451
C_4-COOC	0.421	4.013	0.560	9.887	15.811
C_5-COOC	0.441	4.248	0.562	10.691	14.956
W_d-COOC		4.045	0.500	6.909	12.739
COC-COOC		4.446	0.514	9.328	16.518
NC_3-COOC		4.890	0.532	10.221	15.844
PO_4-PO_4		8.318	0.528	15.665	16.804
C_3-PO_4	0.370	3.995	0.522	13.388	14.145
C_4-PO_4		4.533	0.526	14.159	15.477
C_5-PO_4		4.721	0.526	15.311	14.640
W_d-PO_4		6.197	0.512	9.894	12.469
COC-PO_4		5.544	0.511	13.359	16.168
NC_3-PO_4		6.183	0.488	14.638	15.509
COOC-PO_4		7.012	0.529	10.938	17.167

[a]Parameters taken from ref. 1. [b]Parameters taken from ref. 2.

Table SI-2. Thermodynamic properties of single-component systems.

Molecule	ρ (kg/m^3)	ΔH_{vap} (kJ/mol)	Γ (mN/m)	ΔG_s (kJ/mol)
W_d−W_d	999±4/997	45.9±0.2/44.0	69.4±2.0/72.0	-26.4±1.0/-26.5
C_3	494±5/493	14.9±0.1/14.8	6.5±0.5/6.6	-8.6±0.1/-8.3
C_4	573±4/573	21.2±0.2/21.0	12.1±0.7/11.9	-11.8±0.1/-11.4
C_5	597±4/621	25.1±0.2/26.4	15.5±0.8/15.5	-14.2±0.1/-14.2
C_3−C_4	664±4/682	36.4±0.3/36.6	19.5±0.9/19.8	-20.0±0.1/-19.7
C_3−C_5	662±4/699	40.4±0.4/41.5	22.0±0.7/21.1	-23.0±0.1/-22.2
C_4−C_5	680±4/715	45.5±0.4/46.6	22.4±0.8/22.4	-25.5±0.2/-24.7
COC	652±4/655	18.7±0.1/18.5	11.3±0.7/11.4	-10.1±0.1/-10.2
COC−C_3	704±3/739	32.6±0.3/32.4	19.3±0.9/19.5	-17.8±0.1/-17.5
COC−C_4	723±3/756	38.0±0.3/36.9	21.5±0.8/21.5	-20.5±0.1/-20.6
C_5-COC-C_5	754±3/790	66.9±0.5/63.5	26.9±0.9/25.4	-36.9±0.2/-36.0
NC_3	599±4/629	22.8±0.2/24.0	14.6±0.7/13.4	-12.4±0.1/-11.7
NC_3−C_3	683±3/716	37.0±0.3/34.4	20.0±0.8/19.2	-20.1±0.1/-19.6
NC_3−C_4	702±3/732	39.2±0.3/39.1	20.3±0.8/20.8	-23.6±0.1/-23.2
NC_3−C_5	713±3/744	43.2±0.4/43.8*	21.9±0.8/21.5	-25.8±0.2/-25.8
[a]COOC−C_3	825±3/890	38.2±0.4/40.2	23.7±0.8/24.6	-22.0±0.1/-21.1
[a]COOC−C_4	840±3/883	42.4±0.4/44.1*	24.1±0.9/25.7	-24.3±0.2/-24.3
[a]COOC−C_5	845±3/878	45.4±0.4/48.1*	24.7±0.9/26.6	-26.6±0.2/-26.7
[b]PO_4−C_3	1231±4/1202	52.3±0.4/50.0	32.9±0.8/37.0	-29.0±0.2/-30.4

Properties (left) in green and blue mean the deviation from experiment (right and black)[3-5] is in the range of δ < 5.0% and δ ≥ 5.0%, respectively. *Properties experimentally unavailable are generated by extrapolating from the properties of the corresponding homologues. [a]Reference properties are set as the average of the properties of $COOC_x$ and C_xOOC with x equaling to 4, 5, or 6. [b]Experimental properties[3-5] of trimethyl phosphate are set as reference.

Table SI-3. Properties of binary mixtures.

Cross-term	ρ_{mix} (kg/m^3)	E_{int} (kJ/mol)	f_{pe}	ΔG_{hyd} or ΔG_C (kJ/mol)	Γ_{int} (mN/m)
C_3-W_d	725±3/713	-29.7±0.3/-29.6	1.5%	hyd: 7.7±0.3/8.2	43.6±0.6/48.5
C_4-W_d	752±3/738	-36.6±0.3/-36.7	1.5%	hyd: 8.1±0.3/8.7	47.8±0.3/49.4
C_5-W_d	742±3/753	-42.9±0.4/-43.7	3.6%	hyd: 9.7±0.4/9.8	49.6±0.5/50.0
COC-C_3	551±4/549	-23.2±0.2/-23.1	0.4%		
COC-C_4	594±4/593	-27.6±0.3/-27.1	2.0%	C_{16}: -4.4±0.2/-6.2	
COC-C_5	622±4/626	-28.5±0.3/-29.1	3.9%		
COC-W_d	801±3/870	-54.7±0.5/-57.4	8.8%	hyd: -7.9±0.4/-8.0	
NC_3-C_3	572±4/601	-29.9±0.3/-32.1	5.3%		
NC_3-C_4	602±4/631	-37.1±0.4/-37.0	5.8%	C_{16}: -9.7±0.5/-9.25	
NC_3-C_5	620±4/651	-37.5±0.4/-39.8	3.3%		
NC_3-W_d	896±4/860	-68.6±0.7/-69.8	8.0%	hyd: -12.1±0.6/-13.4	
NC_3-COC	658±4/691	-40.9±0.4/-43.5	3.7%		
COOC-C_3	683±4/715		3.5%		
COOC-C_4	727±3/738		3.3%	C_{16}: -7.6±0.4/-8.3	
COOC-C_5	743±3/752		4.4%		
COOC-W_d	973±5/996		4.4%	hyd: -11.9±0.5/-11.6	
COOC-COC	844±4/847		4.1%		
COOC-NC_3	801±3/839		2.7%		
[a]C_3PO_4-C_3	794±3/807		11.0%	C_6: -21.0±0.9/-24.4	
[a]C_3PO_4-C_4	812±3/825		8.0%	C_7: -19.6±0.8/-23.4	
[a]C_3PO_4-C_5	804±3/832		5.8%	C_8: -19.2±0.9/-22.4*	
[a]C_3PO_4-W_d	1085±5/1133		3.9%	hyd: -35.0±1.5/-36.4	
[a]C_3PO_4-COC	921±5/962		5.1%		
[a]C_3PO_4-NC_3	907±5/933		5.8%		
[a]C_3PO_4-COOC	1099±5/1132		2.9%		

CG Properties (left) in green and blue mean the deviation from the reference data (right) is in the range of δ < 5.0% and δ ≥ 5.0%. Here, the reference ρ_{mix}, E_{int}, and f_{pe} are obtained from AA MD simulations by using the OPLS FF.[6] Free energy and interfacial tension are taken from experimental materials.[7-8] *Properties experimentally unavailable are generated by extrapolating from the properties of the corresponding homologues. [a]The C_3PO_4 group is represented by C_3−PO_4 in CG MD simulations and trimethyl phosphate in the case of AAMD simulations and experiment.

Table SI-4. The bonded parameters of DMPC and DPPC lipids.

Bond	L_b (nm)	k_b (kJ/mol/nm^2)	Angle	θ_a (degree)	k_a (kJ/mol)
		DMPC			
1-2	0.470	1250	1-2-3	140	7.0
2-3	0.390	1250	2-3-4	135	7.0
3-4	0.340	1250	2-3-5	180	7.0
3-5	0.430	1250	3-5-6	180	11.0
5-6	0.520	1250	5-6-7	180	11.0
6-7	0.470	1250	4-8-9	180	11.0
4-8	0.430	1250	8-9-10	180	11.0
8-9	0.520	1250			
9-10	0.470	1250			
		DPPC			
1-2	0.470	1250	1-2-3	140	7.0
2-3	0.390	1250	2-3-4	135	7.0
3-4	0.340	1250	2-3-5	180	7.0
3-5	0.430	1250	3-5-6	180	11.0
5-6	0.520	1250	5-6-7	180	11.0
6-7	0.420	1250	6-7-8	180	11.0
7-8	0.360	1250	4-9-10	180	11.0
4-9	0.430	1250	9-10-11	180	11.0
9-10	0.520	1250	10-11-12	180	11.0
10-11	0.420	1250			
11-12	0.360	1250			

SI.2. Meta-multilinear interpolation parameterization (meta-MIP)

The meta-MIP algorithm[2] was applied here to optimize parameters. It constructs local databases in an on-the-fly pattern, the connection of which leads to an optimal path, thus ensuring a high parameterization efficiency. The basic procedure is introduced as followings.

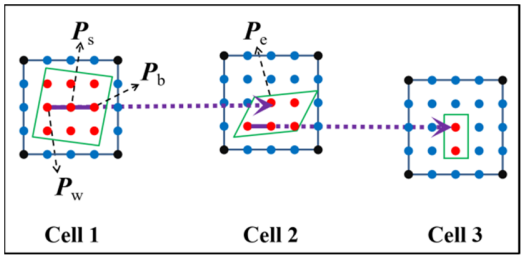

Fig. SI-1. Schematic diagram of meta-MIP.

Step 1. Starting from an initial parameter set $\boldsymbol{P}_s$ (Figure SI-1), Cell 1 encircling $\boldsymbol{P}_s$ is constructed. {ρ, ΔH_{vap}, Γ} at 2^d vertexes (black points) of Cell 1 are evaluated by performing simulations.

Step 2. The resolution inside Cell 1 is raised by multilinear interpolation via

$$f(\boldsymbol{P}_x) = \sum_{n=1,2^d}\left(\left(\prod_{m=1,d} w_{n,m}\right) \times f(\boldsymbol{P}_n)\right), w_{n,m} = \frac{p_{n+1,m}-p_{x,m}}{p_{n+1,m}-p_{n,m}}. \quad \text{(SI-1)}$$

Here, $f(\boldsymbol{P})$ is the property of a d-dimensional $\boldsymbol{P}$ (d=2 in Figure SI-1). The value of $f(\boldsymbol{P}_x)$ is predicted by learning from the 2^d neighbors $\boldsymbol{P}_n$. The weighting factor $w_{n,m}$ stands for the degree of deviation of $\boldsymbol{P}_x$ from $\boldsymbol{P}_{n+1}$. $p_{x,m}$, $p_{n,m}$, and $p_{n+1,m}$ is the component of $\boldsymbol{P}_x$, $\boldsymbol{P}_n$, and $\boldsymbol{P}_{n+1}$ in the mth direction, respectively.

The high-resolution version of database Cell 1 is obtained (blue and red points). At each parameter point, the absolute deviations of {ρ, ΔH_{vap}, Γ} in relative to experimental values are calculated as {δ_1, δ_2, δ_3}.

Step 3. The interpolated parameter sets are screened (red points) by defining a screening condition (SC) condition. For instance, SC={δ_1<200%, δ_2<200%, δ_3<200%} restricts all three properties loosely; SC={δ_1<5%, δ_2<5%, δ_3<200%} reserves accuracy for ρ and ΔH_{vap}; SC={δ_1<5%, δ_2<5%, δ_3<5%} reserves accuracy for all three properties.

Step 4. The screened parameter sets are ranked by defining a ranking condition (RC). For instance, RC={decreasing δ_a=(δ_1+δ_2+δ_3)/3} outputs the parameter sets in the ascending order of δ_a; RC={decreasing δ_3} outputs the parameter sets in the ascending order of δ_3; RC={increasing ε} outputs the parameter sets in the descending order of ε. Afterwards, the worst/best point $\boldsymbol{P}_w$/$\boldsymbol{P}_b$ is obtained as the last/first ranked parameter set. Then, the optimal path in Cell 1 (purple solid line) is roughly obtained as the vector from $\boldsymbol{P}_w$ to $\boldsymbol{P}_b$.

Step 5. The optimization moves to the next cell ($\boldsymbol{P}_e$) by extrapolating the optimal path of Cell 1 (purple dash line). This is reasonable, because the optimal path is roughly linear in local areas, and further screening and ranking could adjust the path direction. An extrapolation factor EF=|$\boldsymbol{P}_e$-$\boldsymbol{P}_b$|/|$\boldsymbol{P}_b$-$\boldsymbol{P}_w$| (suggested value of 0 to 1) is set to control the speed and stability of optimization.

Exit condition. If the predicted properties of $\boldsymbol{P}_b$ from interpolation are satisfying, further MD simulations are performed to test this point. If the simulated properties of $\boldsymbol{P}_b$ (or one of its neighbouring points) are satisfying, the optimization program exits.

Steps 1 to 5 are iterated until an exit criterion is met in Step 5 or an unphysical error (such as crystallization) occurs in any step. A detailed illustration for the usage of meta-MIP could be found in our previous works.[2, 9]

SI.3. Properties of Membranes

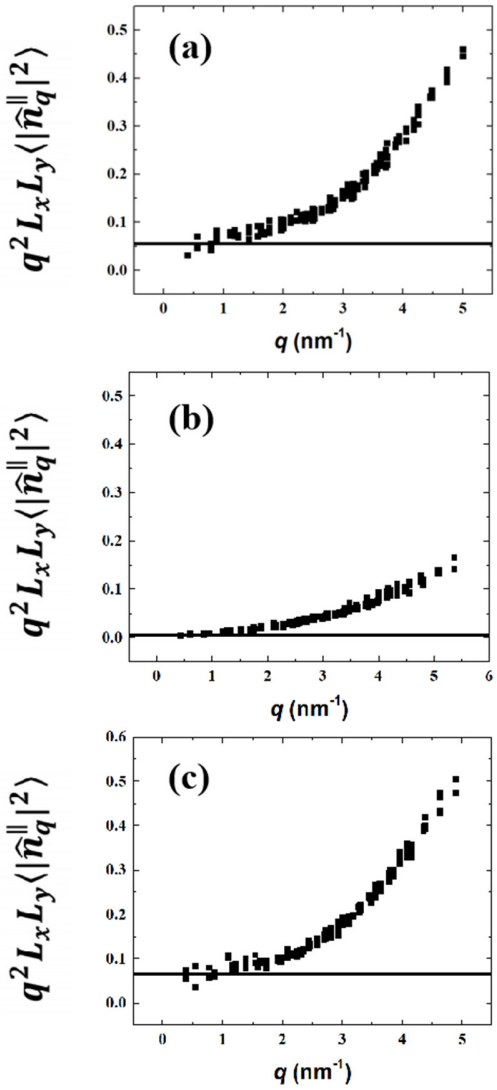

Fig. SI-2. Fluctuation spectra for (a) DMPC membrane at 303 K, (b) DPPC membrane at 303 K, and (c) DPPC membrane at 323 K.

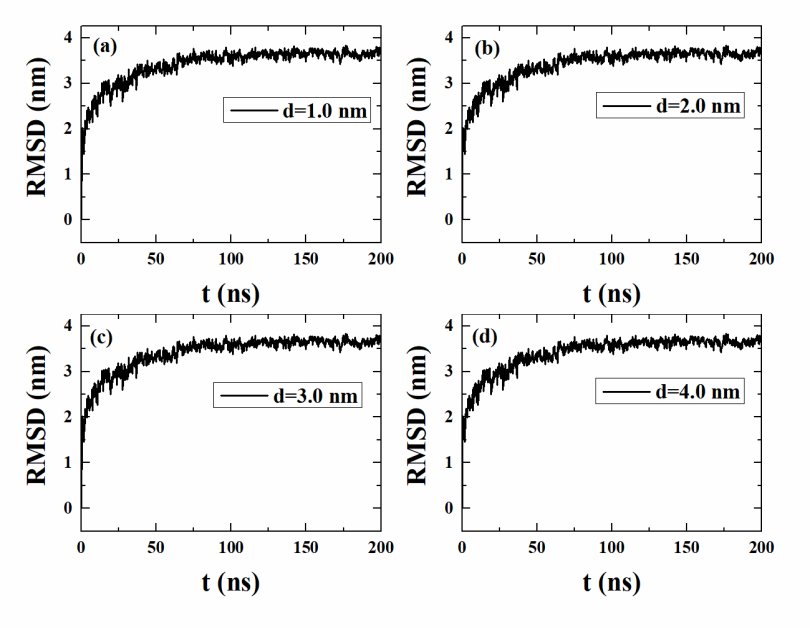

Fig. SI-3. Root mean-square deviation throughout a DMPC membrane at typical umbrella sampling windows where the normal position of a tagged lipid is constrained.

References

1. M. Wan, J. Song, W. Li, L. Gao, W. Fang, Development of Coarse-Grained Force Field by Combining Multilinear Interpolation Technique and Simplex Algorithm. *J. Comput. Chem.* **2020,** *41*, 814-829.
2. M. Wan, J. Song, Y. Yang, L. Gao, W. Fang, Development of coarse-grained force field for alcohols: an efficient meta-multilinear interpolation parameterization algorithm. *Phys. Chem. Chem. Phys.* **2021,** *23*, 1956-1966.
3. J. Rumble, *CRC Handbook of Chemistry and Physics*. 97 ed.; CRC press: Boca Raton, 2017.
4. C. L. Yaws, *Thermophysical properties of chemicals and hydrocarbons*. 2 ed.; William Andrew: Beaumont, 2014.
5. C. L. Yaws, *The Yaws handbook of vapor pressure: Antoine coefficients*. 2 ed.; Gulf Professional Publishing: Houston, 2015.
6. M. J. Robertson, J. Tirado-Rives, W. L. Jorgensen, Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015,** *11* (7), 3499-3509.
7. S. Cabani, P. Gianni, V. Mollica, L. Lepori Group Contributions to the Thermodynamic Properties of Non-Ionic Organic Solutes in Dilute Aqueous Solution. *J. Solution Chem.* **1981,** *10* (8), 563-595.
8. S. Zeppieri, J. Rodríguez, Interfacial Tension of Alkane + Water Systems. *J. Chem. Eng. Data* **2001,** *46* (5), 1086-1088.
9. J. Song, M. Wan, Y. Yang, L. Gao, W. Fang, Development of Accurate Coarse-Grained Force Fields for Weakly Polar Groups by an Indirect Parameterization Strategy. *Phys. Chem. Chem. Phys.* **2021,** *23*, 6763-6774.