Supporting Information For

A Kirkwood-Buff Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20

Elizabeth A. Ploetz, Sadish Karunaweera, Nikolaos Bentenitis, Feng Chen, Shu Dai, Moon B. Gee, Yuanfang Jiao, Myungshim Kang, Nilusha L. Kariyawasam, Nawavi Naleem, Samantha Weerasinghe and Paul E. Smith

Methods

General Simulation

Classical molecular dynamics simulations were performed using Gromacs[1,2] version 2016[25] or 2016.4.[3] Periodic boundary conditions and the minimum image convention were employed. The temperature was fixed at 298 K unless otherwise specified using the Nosé-Hoover[4,5] thermostat with the time constant for coupling groups τ_T = 0.5 ps. For all simulations except the temperature replica exchange[6-10] molecular dynamics (MD) (T-REMD) runs (which were constant *NVT*), isotropic pressure coupling with a reference pressure of 1 bar was achieved using the Parrinello-Rahman[11,12] barostat with the time constant for pressure coupling τ_p = 2.5 ps. All bonds were constrained using the LInear Constraint Solver[13] (LINCS) algorithm for non-water molecules and the SETTLE[14] algorithm for water. SPC/E water was kept rigid. The use of bond constraints allowed for a 2-fs time step to be used for the integration of the equations of motion, which was performed using the Leap-Frog[15] algorithm. The PME technique[16] was used to calculate electrostatic and LJ interactions with a cutoff distance of 0.1 nm for both the real space electrostatic and LJ[17,18] contributions, convergence parameters of 3.123 nm^{-1} (electrostatic Ewald) or 3.351 nm^{-1} (LJ Ewald), cubic interpolation, a maximum fast Fourier transform grid spacing of 0.12 nm for the reciprocal space sum, and tinfoil boundary conditions. We note that geometric combination rules are used in the reciprocal part of the LJ-PME calculation. Because some combination rules are broken in KBFF20, this does introduce very small errors in the forces and energies. The Verlet cut-off scheme[19] was used, with the 'verlet-buffer-tolerance' parameter (which is the allowed energy error due to the Verlet buffer) set to the Gromacs default of 0.005 kJ/mol/ps/atom. If counter-ions were required to create a net neutral system, only enough sodium or chloride was added for neutralization. No bulk salt was used.

Pure Water Simulation and Analysis

With the exception of the excess chemical potential and shear viscosity, the simulated properties shown in Table 1 are the average of 1 μs long simulations of (~4 nm)^3 water boxes. Errors on the last digit are displayed in parentheses and are the standard deviation of ten 100-ns block averages, divided by the square root of ten. For the enthalpy of vaporization, the polarization correction[20] of +5.22 kJ/mol was added to the potential energy, and the enthalpy of vaporization was calculated as $\Delta H_{vap} = -[(PE^{inter} - PE^{intra}) / N + \text{polarization correction}] + RT$, where the intramolecular potential energy (PE) was zero. The static dielectric constant was obtained from 'gmx dipoles', and the self-diffusion coefficient from 'gmx msd'. κ_T was obtained from additional 1 μs simulations at 100, 200, 300, 400, and 500 bar. The system volume *vs*. *p* plot was fit with a quadratic to obtain the κ_T. α_p and C_p were obtained from additional 1 μs simulations at 280, 285, 290, 295, 305, 310, 315, 320, 340, 350, 375, and 400 K. The system volume and enthalpy *vs*. *T* were fit with quadratic equations to obtain α_p and C_p. The excess chemical potential ($\Delta G^*_{solvation}$) was obtained in a two-part process. First, thermodynamic integration was performed to turn the intermolecular Coulomb interactions off for one water (with twenty-one equally spaced windows each simulated for 100 ns with the Coulombic interactions interpolated linearly), giving a

contribution of 38.941(5) kJ/mol. Second, thermodynamic integration was performed to turn the intermolecular LJ interactions off of the "zero-partial charges water" from the previous step (soft core transformation with twenty-one equally spaced windows each simulated for 100 ns), giving a contribution of -9.29(1) kJ/mol. These second TI results gave very similar values to a 100 ns Widom test particle insertion[21] of one LJ-water into pure water, which gave a free energy difference of 9.26 kJ/mol. For the shear viscosity, an additional 20 ns *NVT* simulation was ran in which the off-diagonal elements of the pressure tensor were written out every 2 fs. The Einstein relation[22,23] was then used (taking linear fits between 0 – 100 ps from each of eight 2.5 ns subaverages) to calculate the shear viscosity. The error given is the standard deviation of these subaverages, divided by the square root of eight.

Binary Mixtures

The experimental KB inversion procedure involves finding experimental activity/chemical potential, density/volume, and the solution isothermal compressibility datasets as a function of composition at the desired temperature and pressure for the system of interest. Analytic fits of these properties versus composition are then used to feed the chemical potential derivatives with respect to composition, partial molar volumes, and isothermal compressibility into the KB inversion equations for a given composition to provide the KB integrals. Details for the experimental analyses of the systems described in the KBIs for Mixed Interactions section (besides the previously described &/or published aqueous NMA, NaCl, and MOH systems) are located in Tables S4-S5.

For all systems in which the KBIs were analyzed, the system sizes were (~10 nm)^3 and the simulation time was 100 ns. The last 80 ns were analyzed using four 20 ns block averages. Errors shown are the standard deviation of the four subaverages. The NMA + water simulation and analysis methods are described in the Theory and Background section. For this system, compositions simulated included x_{NMA} = 0.0, 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0. We also note that the simulated self-diffusion coefficient for pure NMA at 313 K was 0.514(7) × 10^{-9} m^2/s compared to the experimental[24] value at 313 K of 0.458 × 10^{-9} m^2/s. The simulated value was obtained from the slope of the molecular mean squared displacement averaged over 100 NMA molecules *vs*. time (Einstein relation as implemented in the 'gmx msd' Gromacs analysis code) from five 10-ns subaverages and was uncorrected for finite system size effects.[25] For more information on the NMA model, see the NMA[26] KBFF publication. For NaCl + water, simulations of 1, 2, 3, and 4 *m* were performed and the KBIs were averaged between 0.85-1.20 nm, consistent with the original NaCl[27] KBFF publication. For methanol (MOH) + water, simulations of x_{MOH} = 0.125, 0.250, 0.375, 0.500, 0.625, 0.750, and 0.875 were performed and the KBIs were averaged between 1.3-1.7 nm, which differs from the original[28] MOH KBFF publication but is consistent with our more recent[29] publication analyzing MOH + water. Details for the simulations of the other systems described in the KBIs for Mixed Interactions section are located in Table S6.

Tripeptides and Disulfide Test System

The Ac-AXA-NHM peptides (Ac = acetyl, NHM = N-H methyl) were simulated in (~4 nm)^3 boxes, where X is each of the 20 standard amino acids excluding proline, for >1 μs ("KBFF fit") or 5 μs ("KBFF fit + shift"). The error bar on the reported properties is the standard deviation on four subaverages. The disulfide test system was the same, except the model system was two Ac-ACA-NHM peptides in which the cysteines were in a disulfide bond.

Generation of 2-D (ϕ,ψ) Correction Maps (CMAPs)

2-D PMFs, PMF(ϕ,ψ), of the target (database) and simulation ϕ,ψ distributions for a given residue X from the above described Ac-AXA-NHM simulations were created from PMF(ϕ,ψ) = $-RT \ln[P(\phi,\psi)]$ using a bin width of 15°. To generate data in the barrier regions where no data was present, Matlab R2017a (9.2.0.556344) was used. A sample Matlab .m file is included in the File S1; the two key steps are as follows. First, the Matlab Central File Exchange 'gridfit' version 1.1.0.0 script of D'Errico[30] was used to turn the input PMF(ϕ,ψ) data points into a 2-D surface. Second, the built-in Matlab 'interp2' function was used to do a 'spline' interpolation (bicubic spline using 'not-a-knot' end conditions) to generate values for those (ϕ,ψ) bins in which data was missing in the input PMF(ϕ,ψ) data. PMF(ϕ,ψ) values that were already present in the input PMF data were not altered during this process; only PMF(ϕ,ψ) bins which had no information in the input file were generated (approximated) using this approach. This method was able to create barrier peaks, which were absent in the input data, based upon the barrier walls that were present in the input data. This was an advantage over other interpolation approaches that would have cut the peaks off (interpolated the PMF values) or behaved poorly (such as creating unphysical oscillations in the PMF). As an example, the input and output database Thr PMFs are shown in Figure S7 (made using the code provided in File S1).

Once the target and simulation PMF(ϕ,ψ) data was processed in this fashion, the CMAP was calculated as CMAP(ϕ,ψ) = [$\mathrm{PMF}_{\mathrm{target}}(\phi,\psi) - \mathrm{PMF}_{\mathrm{MD}}(\phi,\psi)$ + (optional) 1-D Gaussian(s) on ϕ to increase the barriers] and then used in the subsequent simulation. This process was iterated until the CMAP(ϕ,ψ) converged.

The initial simulations for residue X were of the 'basemap', in which the X residue had no potential on ϕ or ψ and no CMAP. However, when sampling was inadequate, a biasing CMAP was applied to enhance sampling.

When creating the $\mathrm{PMF}_{\mathrm{sim}}(\phi,\psi)$, typically (except for Gln) a bin was only given a value if there were ≥5 observations. We found this to be necessary so that the simulated barriers did not appear too high. If the simulations had had good enough sampling, then the choice of requiring five observations instead of one would obviously not have mattered; however, it was observed to matter in our simulations. For Gln, only one observation was required. For the databases, typically a bin was given a $\mathrm{PMF}_{\mathrm{target}}$ value if only one observation was present (except for Gly and Pro, in which ≥ 5 observations were required), which makes the target barriers appear higher. As made clear in this paragraph, some artistic license was taken during this process, unfortunately. Suffice

it to say, the barriers were achieved through the combined use of the 1-D Gaussians placed on ϕ and/or the adjustment of the number of observations required to calculate a PMF(ϕ,ψ).

Pre-proline Tetrapeptides

The Ac-AXPA-NHM peptides were simulated in (~4 nm)^3 boxes, where X is A, Q, P, or G. The regular MD runs and T-REMD runs were performed for 5 μs and 1 μs, respectively, in an effort to ensure adequate sampling of the pre-proline ϕ and ψ 1-D PMFs in the α region. In the T-REMD simulations, the modifications to the peptide bond (ω) potentials described in the Results section was used to ensure that all peptide bonds remained trans. The T-REMD runs consisted of 26 replicas with *T*'s of 298.00, 303.95, 309.98, 316.10, 322.31, 328.61, 334.99, 341.47, 348.04, 354.72, 361.49, 368.36, 375.35, 382.42, 389.60, 396.90, 404.29, 412.04, 419.66, 427.39, 435.25, 443.21, 451.30, 459.49, 467.82, and 470.00 K. These *T*'s were generated using van der Spoel's online server.[31,32]

Omega Torsion Test System

To compare the ω torsional profiles to experiment, a separate simulation of Ac-AQPA-NHM was performed, in which the only changes compared to the above described pre-proline tetrapeptide simulations is that biases were applied to each ω (in the form of changes to the ω torsional potentials) to allow the barrier to rotation around each ω to be overcome (*i.e.*, to allow sampling of cis and trans peptide bonds). The biases included using both (1) a symmetric two-fold periodic proper torsional potential (phase shift = 180°) with a force constant of 10 kJ/mol instead of the KBFF 'ome2' value of 44 kJ/mol, and (2) a one-fold periodic proper torsional potential that destabilizes cis (phase shift = 0°) with a force constant of 0 kJ/mol instead of the KBFF 'ome1' value of 7 kJ/mol. Note that the 'ome1' potential is not applied on proline residues. So, for proline, only the one-fold potential was changed. We used statistical mechanical perturbation theory to predict the unperturbed (*i.e.*, KBFF20) PMFs.

Free Energies of Hydration

Two-step thermodynamic integration was performed for simulations at 298 K and 1 bar of one solute in a (~4 nm)^3 box of SPC/E water. In Step 1, the intermolecular Coulomb interactions were turned off the solute (with twenty-one equally spaced windows with the Coulombic interactions interpolated linearly). In Step 2, the intermolecular LJ interactions were turned off of the zero-partial charges solute (soft core transformation with twenty-one equally spaced windows). V-rescale temperature coupling[33] was used for all but the NPT, IBT, PRP, and ETH simulations, for which Nosé-Hoover was used. Plots of $< \partial E_{pot}(\lambda) / \partial \lambda >_\lambda$ *vs*. λ, where $E_{pot}(\lambda)$ is the potential energy of the system at a particular λ value and λ is the coupling parameter, were splined (cubic) and then integrated from $\lambda = 0$ (fully on) to $\lambda = 1$ (fully off) to give the change in free energy between $\lambda = 0$ and $\lambda = 1$. After taking the sum for both Steps and multiplying by minus one to

reverse the process, the resulting value was compared to the experimental Gibbs energy change associated with the solvation of the small molecule in water as provided by Graziano[34] or Head-Gordon.[35] The values reported in parentheses in Table S10 are the standard deviation based on three to six 10-ns subaverages (NPT, IBT, PRP, ETH), five 2-ns subaverages (MSH, MSM, DDS, MOH, MIN, pCr, ACT, NMA, MIMA/MIMB), or ten 10-ns subaverages (iPr).

Terminal Group Torsional Parameters

For the N-terminal NH_3^+ and NH_2, the H-N-C-C torsions were taken from LysH. Except for N-terminal Glycine, the N-C-C-$N^{(+1)}$ (ψ) torsion was parameterized based on populations instead of free energies to give an ~5-15% population for $-120° \leq \psi \leq 0°$ in a simulation of NH_3^+-Ala-Ala-CO_2^- in water. This simulation lacked sampling in this region before being parameterized. The target data in this case was the N-terminal ψ population of the Dunbrack[36,37] database (N=62) and the Richardson's Top500angles dataset[38] (N=132). For NH_2^+-Pro-… and NH-Pro-… the same parameters were used as for non(Gly) N-termini, but this was not tested for accuracy due to the scarcity of database data for N-terminal prolines. The NH_2^+-Pro-… charge distribution was taken from a study of zwitterionic amino acids.[39] Because there were only seven N-terminal Glycine in the Richardson's Top500angles dataset, the target data in this case was taken from the redundant PDBr (N=54,212 without symmetrizing), to give a free energy difference between extended and folded ψ values of around 6 kJ/mol (with extended being more favorable). For NH_3^+-Gly-… and NH_2-Gly-…, the ψ torsions were parameterized using a simulation of the hairpin GB1 peptide, which has a positively charged N-terminal Gly at neutral pH.

For the C-terminal CO_2^-, the N-C-C-O torsions were taken from the χ_2 of Asp. For the N-C-C-O of C-terminal CO_2H, the parameters were taken from the χ_2 of Asn. For the C-C-O-H torsion of CO_2H, the parameters were taken from AspH. For the C-terminal NH_2 blocking group, the 2-fold on the ω (peptide bond) torsion was used, which provides a barrier to rotation. For the C-terminal NH-methyl blocking group (NHM), the 1-fold and 2-fold from the ω torsions were used, which provides both a barrier to rotation and cis/trans asymmetry.

Supporting Tables, Figures, and Files

Table S1. Dependence of NMA+SPC/E Excess Coordination Numbers on Long-Ranged Potential Truncation Settings. Current simulations used LJ PME and the Verlet cut-off scheme. Prior simulations of the same system size and simulation length used a LJ cut-off of 0.15 nm and the Group cut-off scheme.

$x_{NMA} = 0.1$	N_{ww}	N_{cc}	N_{cw}
Exp[26]	0.2	-0.4	-2.4
Current Sim	0.97(5)	0.06(3)	-4.2(1)
Previous Sim[40]	1.00(7)	0.06(3)	-4.3(1)
$x_{NMA} = 0.2$	N_{ww}	N_{cc}	N_{cw}
Exp[26]	0.4	-0.7	-1.2
Current Sim	1.06(3)	-0.52(1)	-1.91(3)
Previous Sim[40]	1.33(6)	-0.46(1)	-2.17(5)

Table S2. Ion Lennard-Jones 6-12 parameters. The following combination rules were used: $\sigma_{ij}=(\sigma_{ii}\sigma_{jj})^{0.5}$ and $\varepsilon_{ij}=s(\varepsilon_{ii}\varepsilon_{jj})^{0.5}$. The scaling factor, s, was set to unity for all interactions except for cation to SPC/E water oxygen (OW) interactions, which are noted in the table.

Cations					
Ion	σ_{ii} (nm)	ε_{ii} (kJ/mol)	$\varepsilon_{i\text{-OW}}$ (kJ/mol)	Scaling Factor, s	Ref.
Li^+	0.182	0.700	0.2700	0.40	[41]
Na^+	0.245	0.320	0.3420	0.75	[27]
K^+	0.334	0.130	0.2327	0.80	[41]
Rb^+	0.362	0.150	0.2655	0.85	[41]
Cs^+	0.413	0.065*	0.1954	0.95	[41]
Mg^{2+}	0.210	0.750	0.0699	0.10	[42]
Ca^{2+}	0.290	0.470	0.3871	0.70	[42]
Sr^{2+}	0.310	0.500	0.5704	1.00	[42]
Ba^{2+}	0.380	0.200	0.4329	1.20	[42]

*The Cs^+ ε_{ii} value was incorrectly reported as 0.0065 kJ/mol in the original publication.[41]

Anions			
Ions	σ_{ii} (nm)	ε_{ii} (kJ/mol)	Ref.
F^-	0.370	1.000	[41]
Cl^-	0.440	0.470	[27]
Br^-	0.476	0.300	[41]
I^-	0.535	0.200	[41]

Table S3. Databases

Database	Source	Brief Description	Details	Modifications	Used For	Ref.
PDB	Non(Gly,Ala): Dunbrack's 2010 Backbone-dependent Rotamer Library Gly,Ala: PISCES server analysis by M. Feig (personal communication)	Subset of the PDB	Non(Gly,Ala): Distinguishes cis/trans Pro; Distinguishes CysH and Cys2; Gly & Ala not present; Resolution ≤1.8 Å, an R-factor cutoff of 0.22, & mutual sequence identity of the chains of ≤50%.	None (includes pre-Pro)	Comparison only (for most things); Along with 'Top500angle' used as target for N-terminal non(Gly) 1-D ψ population.	36,37,43
Coil	Fitzkee *et al.*'s Protein Coil Library (Pre-compiled list)	Non-α-helix, non-β-sheet fragments derived from the PDB	<20% sequence identity, <1.6 Å resolution, and a refinement factor of 0.25 or better.	Residues before a proline were removed (except for the pre-Proline study).	Comparison only (for most things); Target for pre-Proline CMAPs (1-D ψ PMFs). Used in target for GlyPro CMAP.	44,45
Coil/NT	Same as above (SAA)	SAA	SAA	Residues before a proline or in turns were removed	Target for simulated Ramachandran plots (CMAPs) and for 1-D χ PMFs.	44,45
PDBr	Rsync (RsyncPDB.sh) of PDB database (rsync.wwpdb.org, as of 1/23/18).	Full rsync of the PDB.	Rsync of PDB without further refinement.	None	Used in target for GlyPro CMAP in regions where the Coil was missing data. Used as target for N-terminal Glycine ψ torsion.	46
Top500angle	Richardson's Top500 Angles Dataset	500 high quality structures from the PDB	≤1.8 Å resolution and low homology structures; high B-factors removed.	None	Along with 'PDB', used as target for N-terminal non(Gly) 1-D ψ population.	38

Table S4. Pure Component Properties for "Mixed Interaction" KB Inversion Procedures

System (2+1)	T(K)	MW_2 (g/mol)	Pure 2		Pure 1		κ_T
			Vol. (cm^3/mol) / dens. (g/cm^3) / phase	Ref	Vol. (cm^3/mol) / d_0 (g/cm^3) / phase	Ref	
$MeNH_3Cl$ +H_2O	298	67.518	55.526 / 1.216 / (*s*)	[47]	18.068 / 0.99705 / (*l*)	[48,49]	*
NaOAc +H_2O	298	82.034	54.255 / 1.512 / (*s*)	[50]	18.068 / 0.99705 / (*l*)	[48,49]	*
β-Ala +H_2O	298	89.093	62.395 / 1.428 / (*s*)	[51]	18.068 / 0.99705 / (*l*)	[48,49]	*
GlyGly +H_2O	298	132.1196	87.150 / 1.516 / (*s*)	[52]	18.068 / 0.99705 / (*l*)	[48,49]	*
GdmCl +H_2O	298	95.533	69.985 / 1.365 / (*s*)	[53]	18.068 / 0.99705 / (*l*)	[48,49,54]	*
GdmOAc +H_2O	298	119.12	97.559 / 1.221 / (*s*)	[55]	18.068 / 0.99705 / (*l*)	[48,49]	*
$GdmEtCO_2$+ H_2O	298	133.16	104.113 / 1.279 / (*s*)	[56]	18.068 / 0.99705 / (*l*)	[48,49]	*
HEACl +H_2O	298	97.55	70.688 / 1.38 / (*s*)	[57]	18.068 / 0.99705 / (*l*)	[48,49]	*
HEAA +H_2O	340	121.137	107.144 / 1.131 / (*l*)	[58]	18.456 / 0.97945 / (*l*)	[48,49]	$\kappa_{T,1}^o$=4.49E-5 bar^{-1} (Ref [48,49]) $\kappa_{T,2}^o$=2.96E-5 bar^{-1} (at 333K, which may be the subcooled liquid)[59]
NaOAc +MOH	298	82.0343	54.255 /1.512 / (*s*)	[50]	40.749 / 0.78633 / (*l*)	[60]	$\kappa_{T,1}^o$=10.5E-5 bar^{-1} (Ref[60])
NMA +MOH	313	73.09	77.648 / 0.9413 / (*l*)	[61]	41.494 / 0.7722 / (*l*)	[61]	$\kappa_{T,1}^o$=1.435E-4 bar^{-1} (at 318K)[62] $\kappa_{T,2}^o \approx \kappa_{S,2}^o$ =5.9553E-5 bar^{-1} (Ref [63])
NaOAc +NMA	304	82.0343	54.255 / 1.512 / (*s*)	[50]	76.832 / 0.951295 / (*l*)	[64](303K)	$\kappa_{T,1}^o \approx \kappa_{S,1}^o$ = 5.63E-5 bar^{-1} (303K, Table S3)[64]
KOAc +NMA	304	98.15	63.528 / 1.545 / (*s*)	[65]	76.832 / 0.951295 / (*l*)	[64](303K)	$\kappa_{T,1}^o \approx \kappa_{S,1}^o$ = 5.63E-5 bar^{-1} (303K, Table S3)[64]

MW (g/mol) of component 1: H_2O=18.015, MOH=32.042, NMA=73.09.

*Unless otherwise noted, $\kappa_{T,1}^o = 4.53\times10^{-5}$ bar^{-1} (Ref [48,49]) and $\kappa_{T,2}^o$ approximated as zero.

Table S5. Mixture Properties for "Mixed Interaction" KB Inversion Procedures

System (2+1)	T(K)	Volumetric Data		Activity Data	
		Fit	Raw Data	Fit	Raw Data
$MeNH_3Cl+H_2O$	298	d (g/cm^3) = $d_0 + a\rho_2 + b\rho_2^2$, a = 1.3420065E-2, b = -1.7383354E-3	[66]	$\ln\gamma_\pm = -\ln(10)A_\gamma m_2^{0.5}/(1+bm_2^{0.5}) - \ln(1-m_2 g_w)$, A_γ = 0.5115, b = 0.9576242, g_w = 5.4119743E-2	[67]
$NaOAc+H_2O$	298	d (g/cm^3) = $d_0 + (a+bt+ct^2)\rho_2 + (d+et)\rho_2^{1.5}$, t = 25 deg C, a = 4.364e-2, b = -6.74e-5, c = 6.482e-7, d = -2.113e-3, e = -9.843e-6	Fit was taken from Söhnel, except a slightly different value of d_0 was used by Söhnel.[68]	$\ln\gamma_\pm = -\ln(10)am_2^{0.5}/(1+bm_2^{0.5}) + cm_2$, a = 0.4650, b = 1.579202, c = 1.76298	[69]
β-Ala+H_2O	298	d (g/cm^3) = $d_0 + am_2 + bm_2^2 + cm_2^3$, a = 3.0136697003193684E-2 b = -2.2008402079345890E-3 c = 8.2409776239040120E-5	Only raw data from Ref [70] was used, not fit.	$\ln\gamma_2 = am_2 + bm_2^2 + cm_2^3 + dm_2^4$, a = -3.6750860135339503e-2, b = 2.8103884298331403e-2 c = -3.0429162146024371e-3, d = 1.2165306559958484e-4	[71]
GlyGly+H_2O	298	d (g/cm^3) = $d_0 + am_2 + bm_2^2 + cm_2^3$, a = 5.54571301e-2, b = -5.43306535e-3, c = 1.36817034e-4	[72]	$\ln\gamma_2 = am_2 + bm_2^2 + cm_2^3$, a = -0.491913617, b = 0.251205534, c = -5.50714396e-2	[73]
GdmCl+H_2O	298	d (g/cm^3) = $d_0(1 + aw_2 + bw_2^2)$ a = 0.2710, b = 0.033	Fit & coefficients provided in Ref[74] (Same as used in Ref [54])	$\ln\gamma_\pm = -\ln(10)A_\gamma m_2^{0.5}/(1+bm_2^{0.5})$ $+\ln\{0.25c[-1+(1+8m_2/c)^{0.5}]/m_2\}$, A_γ = 0.5115, b = 2.681, c = 3.548	[74] (Same as used in Ref [54])
GdmOAc+H_2O	298	PMV's fixed at pure values	N/A	Pitzer[75] equation with A_Φ = 0.3921, b = 1.2, α = 2.0, β_0 = -4.95727267e-3, β_1= 0.253582090, C_Φ = 1.40791817e-3	[76]
GdmEtCO$_2$+H_2O	298	PMV's fixed at pure values	N/A	Pitzer[75] equation with A_Φ = 0.3921, b = 1.2, α = 2.0, β_0 =-1.65746696e-2, β_1=0.183785945, C_Φ =8.74772202e-4	[77]
*HEACl+H_2O	298	PMV's fixed at pure values	N/A	$\ln a_{water} = \ln(P_{vap}/P_{vap}^0) + B_T(P_{vap} - P_{vap}^0)/RT$ $\log_{10}P_{vap}$(kPa) = $A(m)+B(m)/T+C(m)/T^2$ $A(m)=A_0+A_1m_2+A_2m_2^2+A_3m_2^3$, $B(m)=B_0+B_1m_2+B_2m_2^2+B_3m_2^3$ $C(m)=C_0+C_1m_2+C_2m_2^2+C_3m_2^3$ P_{vap}^0=3.17 kPa, T=298.15 K, B_T=-1193.77 cm^3/mol A_0=10.14146, A_1=-2.783560, A_2=1.012594, A_3=-0.072459, B_0=-3562.851, B_1=1734.523, B_2=-638.4133, B_3=45.46873, C_0=205109.1, C_1=-270944.7, C_2=100403.8, C_3=-7120.746	All parameters including P_{vap}^0 from Ref [78], except B_T, which was from Ref [79].
*HEAA+H_2O	340	V^{ex} (cm^3/mol) = $x_2(1-x_2)[a+b(1-2x_2)+c(1-2x_2)^2+d(1-2x_2)^3]$ a = -2.87275505, b = -2.53286552 c = -1.20110857, d = 0.618181169	Exp data is at 343 K[58]	NRTL equation with $\tau_{HEAA,water}$=(-465.72452 K+0.030143t)/t, $\tau_{water,HEAA}$=(167.397543 K-0.020558t)/t, α=0.9, t=340.15 K	Reference has NRTL parameters & equation[58]
NaOAc+MOH	298	PMV's fixed at pure values	N/A	Pitzer equation with A_Φ = 1.294, b = 3.2, α = 0.0, β_0 =0.127208, β_1=0.0, C_Φ =-0.0256465 (did not use reference's Pitzer parameters)	[80]
**NMA+MOH	313	V^{ex} (cm^3/mol) = $x_1x_2\sum_{i=1}^{4} K_i(2x_2-1)^{i-1}$ K_1 = -1.70507, K_2 = 0.58771, K_3 = -0.23019, K_4 = 0.08042	[61]	NRTL equation with $\tau_{MOH,NMA}$ = -1.18172, $\tau_{NMA,MOH}$ = 1.19768, α = 0.3	Reference has NRTL parameters[81]
NaOAc+NMA	304	PMV's fixed at pure values	N/A	$\ln\gamma_\pm = -Am_2^{0.5}/(1+m_2^{0.5}) + Bm_2 + Cm_2^{1.5}$, A = 0.14128, B = -0.1668, C = -0.0251	Ref. has fit, but sign of 1st term must be changed.[82]
KOAc+NMA	304	PMV's fixed at pure values	N/A	$\ln\gamma_\pm = -Am_2^{0.5}/(1+m_2^{0.5}) + Bm_2 + Cm_2^{1.5}$, A = 0.14128, B = 0.0466, C = -0.0422	Ref. has fit, but sign of 1st term must be changed.[82]

*The activity coefficient models for these salt systems were not electrolyte models. The equations used do not have the Debye-Hückel limiting behavior built into the model and thus should be used with caution. We assumed they were acceptable at $m_2 > \sim 0.5$ mol/kg.

**$G_{NMA\text{-}NMA}$ agrees well with the values given in Ref [83].

Solvent=1, Solute=2. For the ionic systems, in Table S5, the solute (2) refers to the salt not the indistinguishable ion.

d_0 = density of pure solvent in grams/cm^3. ρ_2 = molarity of salt (mol/L).

w_2 = salt mass fraction.

PMV = partial molar volumes.

κ_T treatment was: $\kappa_T = \phi_2^{id}\kappa_{T,2}^o + \phi_1^{id}\kappa_{T,1}^o$. $\phi_2^{id} = x_2V_2^o/(x_1V_1^o + x_2V_2^o)$ and $\phi_1^{id} = 1 - \phi_2^{id}$.

Table S6. Simulated Systems for "Mixed Interaction" Studies

System (2+1)	T(K) (MD)	Compositions Simulated	$G_{22}(R)$ Averaging Region (nm)
$MeNH_3Cl+H_2O$	300	1, 2, 4 m_{salt}	1.5-2.0
$NaOAc+H_2O$	300	1, 3, 5, m_{salt}	1.5-2.0
β-Ala+H_2O	300	1, 3, 5 m_2	1.5-2.0
GlyGly+H_2O	300	0.3, 0.5, 1.0, 1.5 m_2	1.5-2.0
GdmCl+H_2O	300	2, 4 m_{salt}	1.1-1.45 [54]
GdmOAc+H_2O	300	1, 2, 4, 6 m_{salt}	1.5-2.0
$GdmEtCO_2+H_2O$	300	1, 2 m_{salt}	1.5-2.0
HEACl+H_2O	298	0.5, 1.0, 1.5, 2.0 m_{salt}	1.5-2.0
HEAA+H_2O	340	1 m_{salt}, x_2 = 0.2, 0.4, 0.6, 0.8	1.5-2.0
NaOAc+MOH	298	0.5, 1.0, 1.5 m_{salt}	1.5-2.0
NMA+MOH	313	x_2 = 0.1, 0.2, 0.4, 0.6, 0.8	2.0-2.5
NaOAc+NMA	304	0.3, 0.4, 0.5 m_{salt}	2.0-2.5
KOAc+NMA	304	0.3, 0.5, 0.8 m_{salt}	2.0-2.5

All simulations were performed for 100 ns in L ~10 nm cubic boxes.

Table S7. Gromacs individual molecule topology (.ITP) files for Mixed Interactions section
ITP files for $MeNH_3^+$, OAc^-, β-Ala, $EtCO_2^-$, and HEA are below. Other small molecules simulated in this work have already been published in the peer-reviewed literature and so are not provided here.

$MeNH_3^+$ (Methylammonium):

```
[ moleculetype ]
; name  nrexcl
MAM          3

[ atoms ]
;   nr    type   resnr  residu    atom    cgnr  charge   mass
     1      NT       1    MAM      N       1    0.500   14.0067
     2      H        1    MAM      H1      1    0.000    1.0080
     3      H        1    MAM      H2      1    0.000    1.0080
     4      H        1    MAM      H3      1    0.000    1.0080
     5      CH3      1    MAM      C       1    0.500   15.0350

[ bonds ]
;  ai    aj funct           c0           c1
    1     2     2 gb_2
    1     3     2 gb_2
    1     4     2 gb_2
    1     5     2 gb_21

[ angles ]
;  ai    aj    ak funct           c0           c1
    2     1     3     2 ga_10
    2     1     4     2 ga_10
    3     1     4     2 ga_10
    2     1     5     2 ga_11
    3     1     5     2 ga_11
    4     1     5     2 ga_11
```

OAc^- (Acetate):

```
[ moleculetype ]
; name  nrexcl
OAc-    3

[ atoms ]
;   nr    type   resnr  residu    atom    cgnr  charge   mass
     1      CH3      1    OAc-      CH3     1    0.000   15.0350
     2      C2       1    OAc-      CC      2    0.300   12.0110
     3      OT       1    OAc-      O1      2   -0.650   15.9994
     4      OT       1    OAc-      O2      2   -0.650   15.9994

[ bonds ]
;  ai    aj funct           c0           c1
    2     1     2 gb_27
    2     3     2 gb_6
    2     4     2 gb_6

[ angles ]
;  ai    aj    ak funct           c0           c1
    1     2     3  2   ga_22
    1     2     4  2   ga_22
    3     2     4  2   ga_38

[ dihedrals ]
;  ai    aj    ak    al funct           c0           c1
    2     1     3     4     2     iplane2
```

β-Ala (β-Alanine): The H-N-C-C and C-C-C-O torsions were taken from LysH and Glu, respectively. The N-C-C-C torsion was parameterized using an NMR study[84] of the β-Ala %gauche in D_2O.

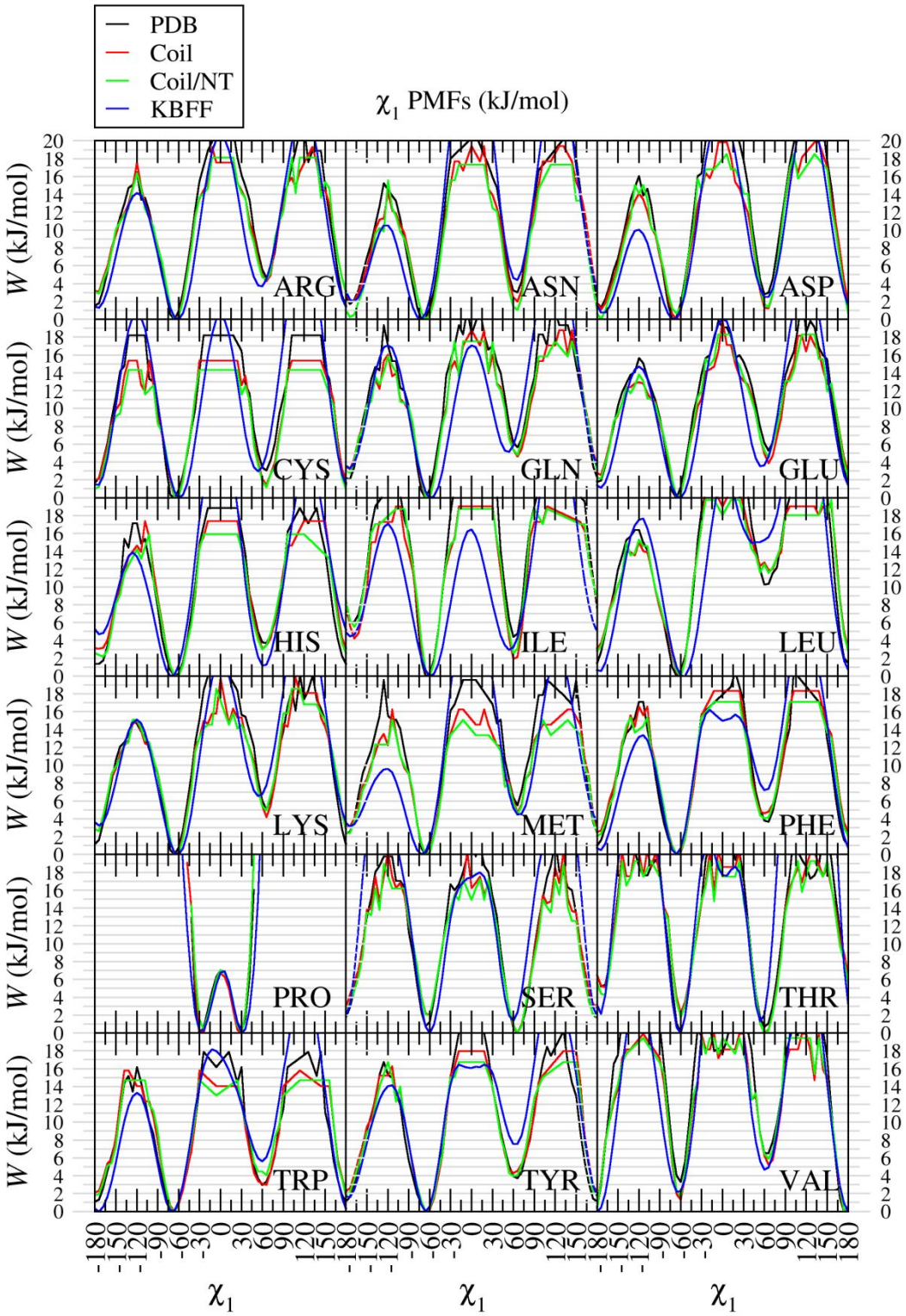
```
[ moleculetype ]
; name  nrexcl
BAL          3

[ atoms ]
;   nr    type   resnr  residu    atom    cgnr  charge   mass
     1      H        1   BAL        H1      1    0.000    1.0080
     2      H        1   BAL        H2      1    0.000    1.0080
     3      NT       1   BAL        N       1    0.500   14.0067
     4      H        1   BAL        H3      1    0.000    1.0080
     5      CH2      1   BAL        CA      1    0.500   14.0270
     6      CH2      1   BAL        CB      2    0.000   14.0270
     7      C2       1   BAL        C       3    0.300   12.0110
     8      OT       1   BAL        O1      3   -0.650   15.9994
     9      OT       1   BAL        O2      3   -0.650   15.9994

[ pairs ]
    1     6
    2     6
    4     6
    3     7
    5     8
    5     9

[ bonds ]
;  ai    aj funct           c0           c1
    3     1     2 gb_2
    3     2     2 gb_2
    3     4     2 gb_2
    3     5     2 gb_21
    5     6     2 gb_27
    6     7     2 gb_27
    7     8     2 gb_6
    7     9     2 gb_6

[ angles ]
;  ai    aj    ak funct           c0           c1
    1     3     2     2 ga_10
    1     3     4     2 ga_10
    1     3     5     2 ga_11
    2     3     4     2 ga_10
    2     3     5     2 ga_11
    4     3     5     2 ga_11
    3     5     6     2 ga_15
    5     6     7     2 ga_15
    6     7     8     2 ga_22
    6     7     9     2 ga_22
    8     7     9     2 ga_38

[ dihedrals ]
;  ai    aj    ak    al funct           c0           c1           c2
    1     3     5     6     9 0.00   1.75   3
    2     3     5     6     9 0.00   1.75   3
    4     3     5     6     9 0.00   1.75   3
    3     5     6     7     9 0.00  -2.364  1
    3     5     6     7     9 0.00  -0.901  2
    3     5     6     7     9 0.00   5.430  3
    5     6     7     8     9 0.00  -6.000  2
    3     5     6     7     9 180.0 -4.500  1
    3     5     6     7     9 0.00   0.000  2
    3     5     6     7     9 0.00   5.430  3

[ dihedrals ]
;  ai    aj    ak    al funct           c0           c1
    7     6     8     9     2 iplane2
```

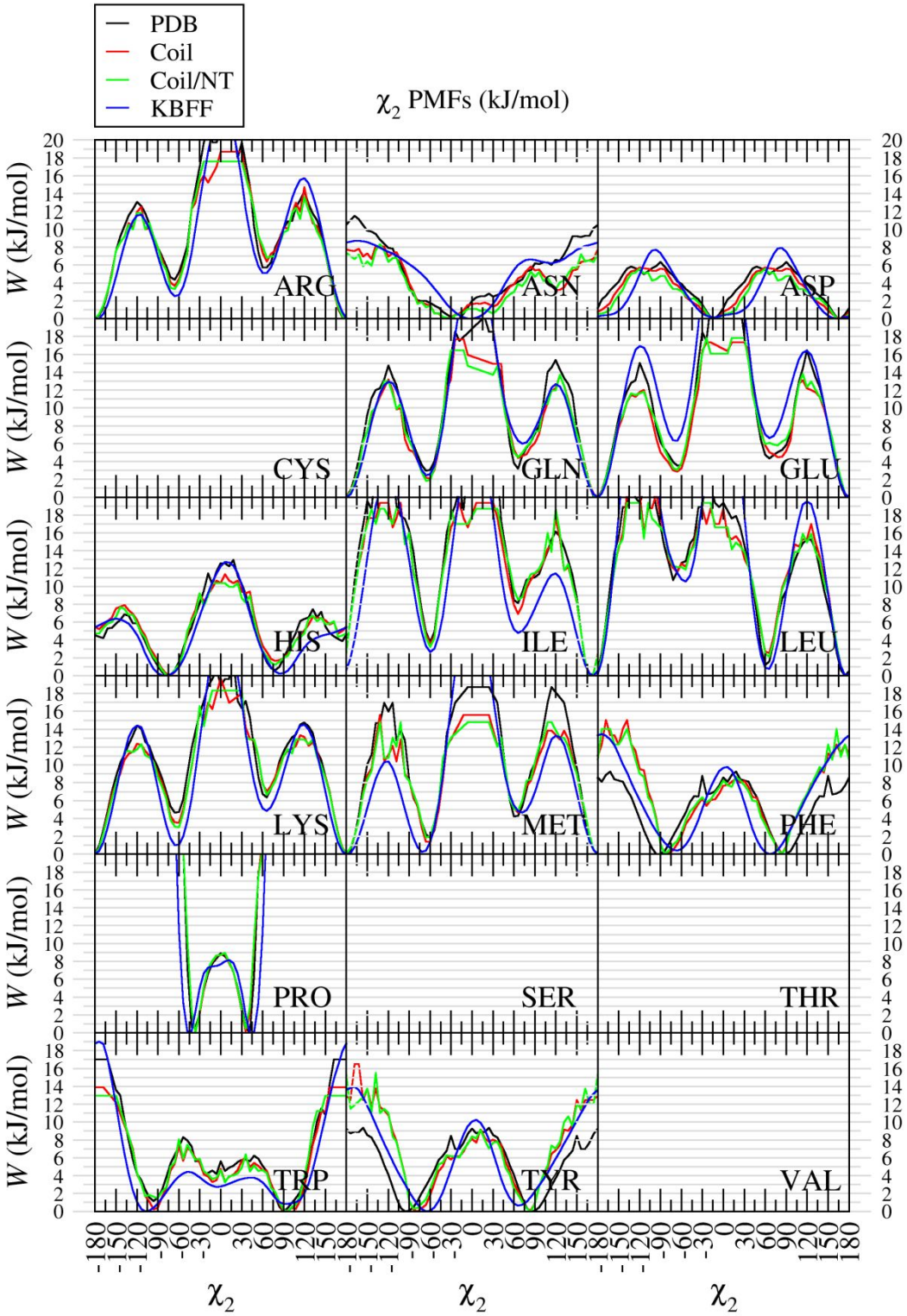
$EtCO_2^-$ (Propionate):

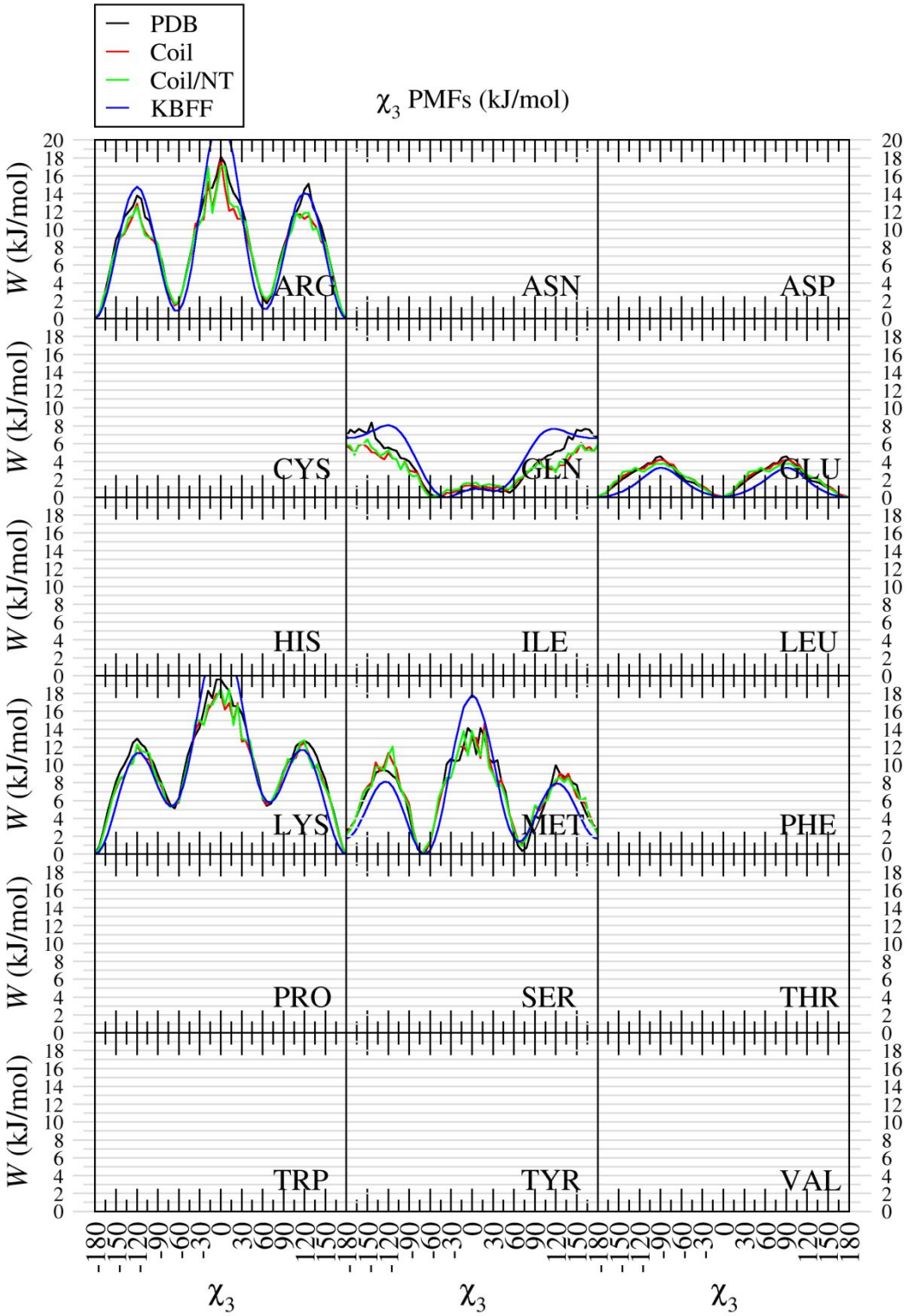
```
[ moleculetype ]
; name  nrexcl
PRP-    3

[ atoms ]
;   nr    type   resnr  residu    atom    cgnr  charge   mass
     1      CH3      1    PRP-      CH3     1    0.000   15.0350
     2      CH2      1    PRP-      CH2     1    0.000   14.0270
     3      C2       1    PRP-      CC      2    0.300   12.0110
     4      OT       1    PRP-      O1      2   -0.650   15.9994
     5      OT       1    PRP-      O2      2   -0.650   15.9994

[ pairs ]
   1  4 1
   1  5 1

[ bonds ]
;  ai    aj funct           c0           c1
   1     2  2   gb_27
   2     3  2   gb_27
   3     4  2   gb_6
   3     5  2   gb_6

[ angles ]
;  ai    aj    ak funct           c0           c1
   1     2     3   2   ga_15
   2     3     4   2   ga_22
   2     3     5   2   ga_22
   4     3     5   2   ga_38

[ dihedrals ]
;  ai    aj    ak    al funct           c0           c1
   3     4     5     2  2   iplane2

 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   1     2     3     4    9           0.00  -6.000  2
```

HEA (2-Hydroxyethylammonium a.k.a. 2-aminoethanol): The H-N-C-C and H-O-C-C torsions were taken from LysH and Ser, respectively. The N-C-C-O torsion was parameterized using an NMR study[85] of the HEACl %gauche values in D_2O.

```
[ moleculetype ]
; name  nrexcl
HEA          3

[ atoms ]
;   nr    type   resnr  residu    atom    cgnr  charge   mass
     1      NT       1    HEA       N       1    0.500   14.0067
     2       H       1    HEA      H1       1    0.000    1.0080
     3       H       1    HEA      H2       1    0.000    1.0080
     4       H       1    HEA      H3       1    0.000    1.0080
     5     CH2       1    HEA      C1       1    0.500   14.0270
     6     CH2       1    HEA      C2       2    0.300   14.0270
     7      O3       1    HEA      O1       2   -0.820   15.9994
     8       H       1    HEA      H4       2    0.520    1.0080

[ pairs ]
    2     6     1
    3     6     1
    4     6     1
    1     7     1
    5     8     1

[ bonds ]
;  ai    aj funct           c0           c1
    1     2     2 gb_2
    1     3     2 gb_2
    1     4     2 gb_2
    1     5     2 gb_21
    5     6     2 gb_27
    6     7     2 gb_18
    7     8     2 gb_1a

[ angles ]
;  ai    aj    ak funct           c0           c1
    2     1     3     2 ga_10
    2     1     4     2 ga_10
    3     1     4     2 ga_10
    2     1     5     2 ga_11
    3     1     5     2 ga_11
    4     1     5     2 ga_11
    1     5     6     2 ga_13
    5     6     7     2 ga_13
    6     7     8     2 ga_12

[ dihedrals ]
;  ai    aj    ak    al   gromos type
   6    5    1    2    9   0.00   1.75   3
   6    5    1    3    9   0.00   1.75   3
   6    5    1    4    9   0.00   1.75   3
   1    5    6    7    9   0.00  -10.0   1
   1    5    6    7    9   0.00   0.5    2
   1    5    6    7    9   0.00   8.5    3
   5    6    7    8    9   0.00   2.480  1
   5    6    7    8    9   0.00   0.500  2
   5    6    7    8    9   0.00   3.000  3
```


PDB
Coil
Coil/NT
KBFF
χ1 PMFs (kJ/mol)
W (kJ/mol)
ARG
ASN
ASP
CYS
GLN
GLU
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
χ1


PDB
Coil
Coil/NT
KBFF
χ2 PMFs (kJ/mol)
W (kJ/mol)
ARG
ASN
ASP
CYS
GLN
GLU
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
χ2


PDB
Coil
Coil/NT
KBFF
χ3 PMFs (kJ/mol)
W (kJ/mol)
ARG
ASN
ASP
CYS
GLN
GLU
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
χ3

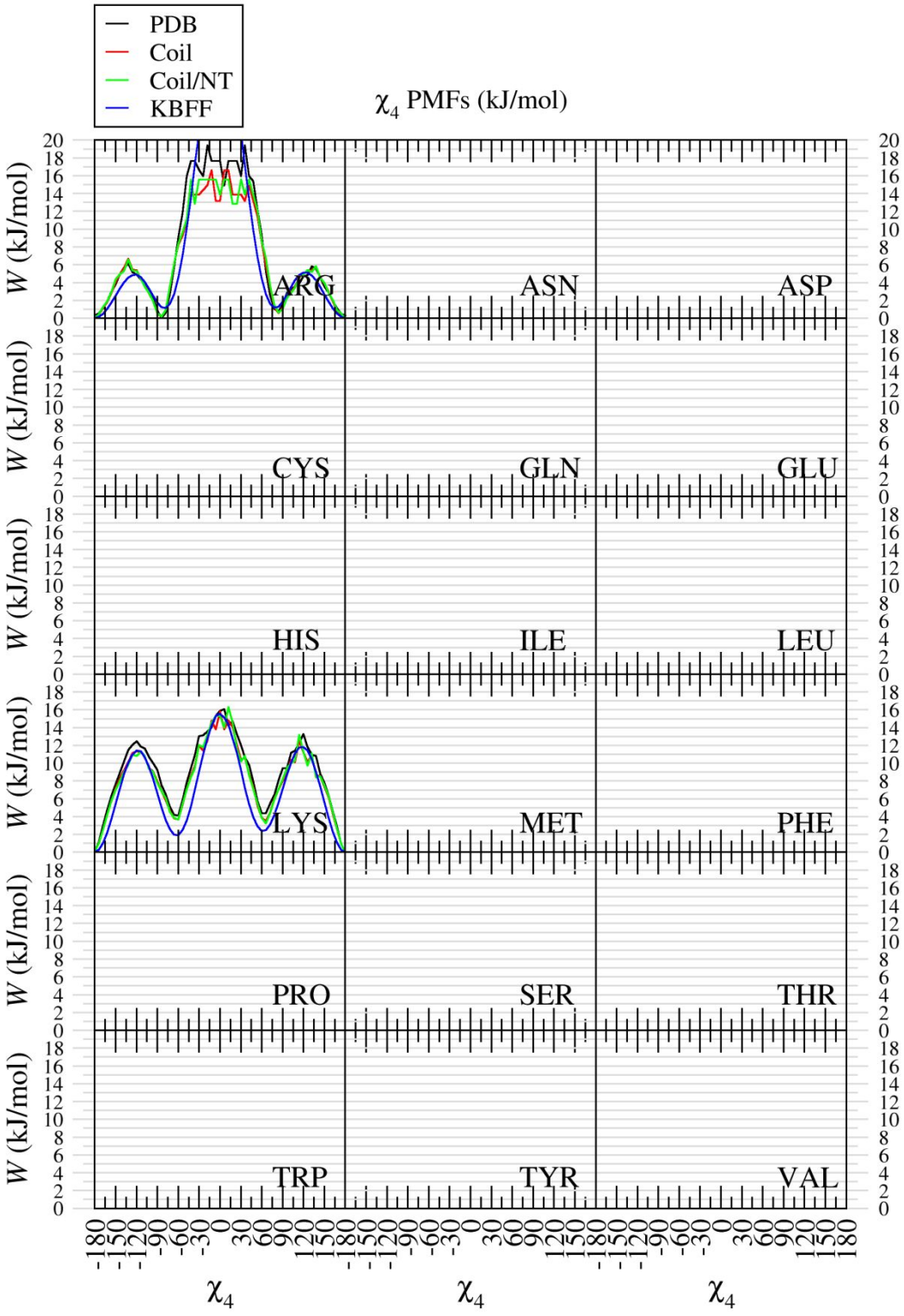


Figure S1. PMFs, *W* (kJ/mol), of χ side-chain dihedral angles from databases[36,37,44,45] *vs.* KBFF20. For titratable residues, the simulation refers to the charged residue, and the database to a mixture of charged and uncharged residues. For Cysteine, the simulation refers to CysH, and the database to a mixture of CysH and Cys2 (except for the PDB, where it is only CysH).

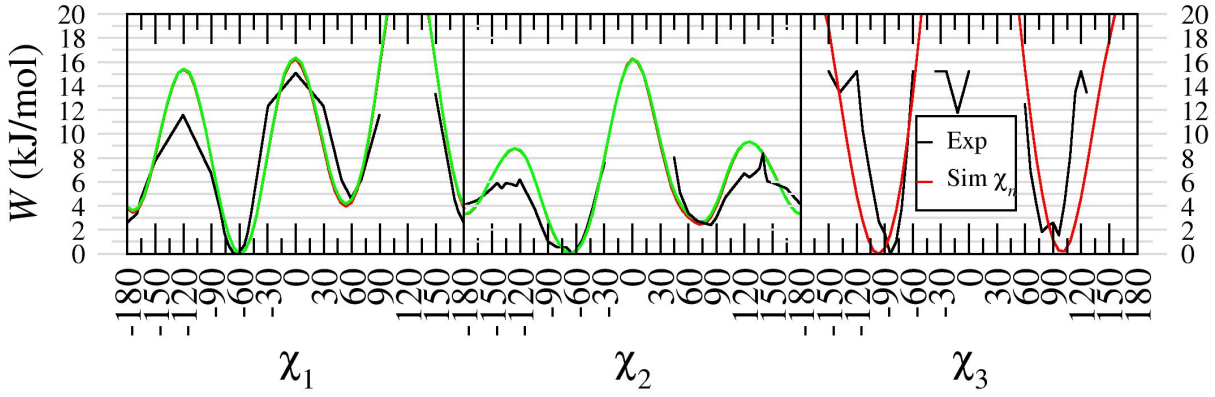


Figure S2. Disulfide χ PMFs, *W* (kJ/mol). Ac-C-NHM Ac-C-NHM was simulated for 5 μs in a (4 nm)^3 box of water (red and green correspond to the two different Cys residues). The experimental data was obtained from a digitization of Figure 1 of Ref [86]. The simulated barrier across $\chi_3 = 0°$ is estimated to be between 30-35 kJ/mol, while the simulated barrier across $\chi_3 = \pm 180°$ is approximately 24.5 kJ/mol.

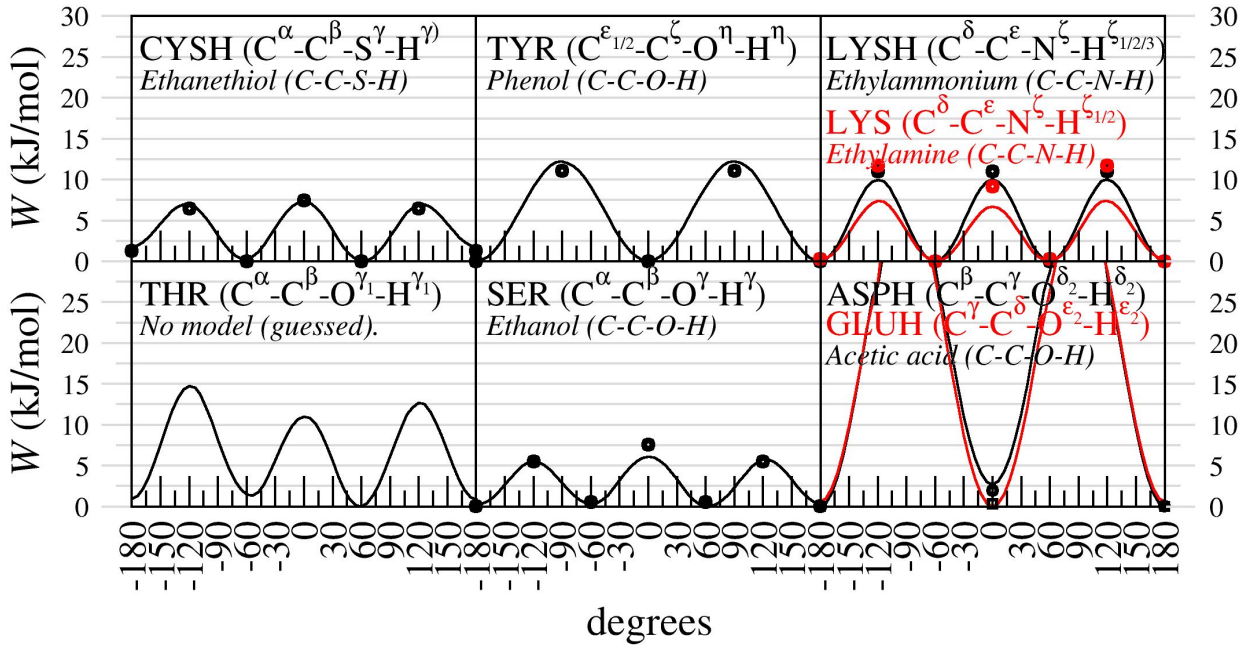


Figure S3. PMFs, *W* (kJ/mol), of side-chain dihedrals for which there is no experimental data compared to small molecule model compound (italics) QM values. Except for acetic acid, the points correspond to QM values taken from Jorgensen *et al.*[87] for small molecule model compounds (italics under residue). The acetic acid points were taken from Terhorst and Jorgensen[88] and correspond to either QM/TIP4P (circles) or QM/RISM (squares). The KBFF20 values (from Ac-AXA-NHM simulations) are shown as curves. For AspH and GluH, KBFF20 barrier heights at ± 90° were not fully sampled but are estimated to be ~40 kJ/mol for AspH and ~33 kJ/mol for GluH. For Lys, the KBFF20 barriers are low because the same C-C-N-H torsional potential is used as on LysH due to our adoption of the additivity philosophy in the creation of KBFF20. Given that there is one fewer C-C-N-H torsion on Lys, the simulated barrier height is ~2/3 what it is for LysH. In order to keep KBFF20 additive in nature, we did not correct the Lys torsion to match the QM barrier height.

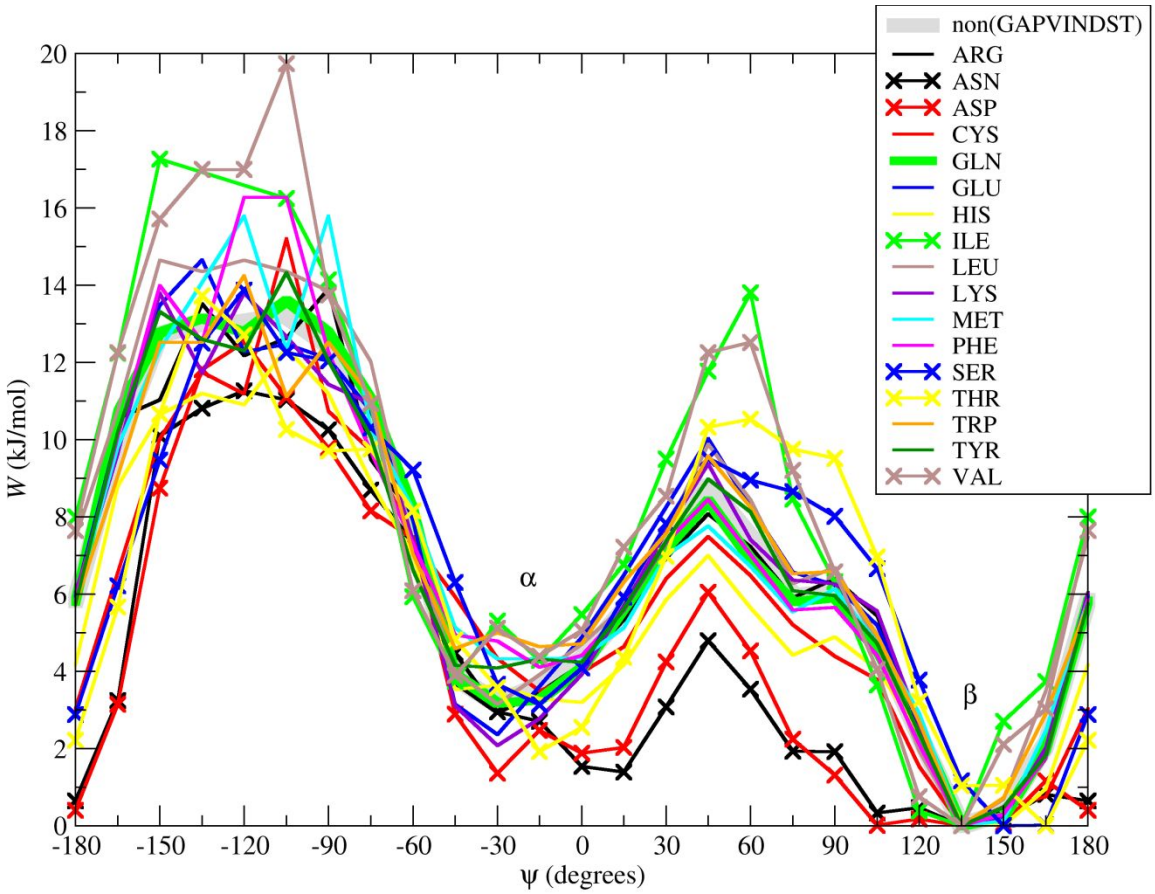


Figure S4. Classification of amino acids. Restricted 1-D PMF, *W* (kJ/mol), profiles for the Coil/NT database[44,45] (using an effective temperature of 300 K). ψ values were evaluated on the condition that ϕ was < 0° or > 120°. For the creation of this figure, only one observation was required to generate a value for a given bin. Gly, Ala, and Pro were not plotted. VINDST residues are shown with the X symbol. They were deemed significantly different from the general case. Gln was chosen to have a profile most representative of the non(GAPVINDST) case. The general vicinity of the α and β minima are labeled schematically on the plot.

Table S8. Summary of 2-D ϕ, ψ CMAPs

Residues the CMAP Applies To	Database Version Used[a]	Atom types that select this CMAP	X (from Ac-AXA-NHM) simulated to create CMAP	Target Database Map	1-D Gaussian barriers along ϕ added to CMAP		
					Height (kJ/mol)	Mean	Std.Dev.
Ala	Coil/NT	CH1A	X=Ala	Ala(N=11,060)	5	0°	30°
					5	135°	40°
All non(GAPNDTC), non(Pre-Pro)	Coil/NT	CH1	X=Gln	Non(GAPVINDST) (N=64,996)	5	135°	20°
Gly	Coil/NT	CH2	X=Gly (symmetrized)	Gly(N=50,804 & then symmetrized)	None		
Pro	Coil/NT	CH1P & N2P	Pro	Pro(N=16,583)	None		
Asp/AspH & Asn	Coil/NT	CH1l	Merge of PMFs from X=Asp & X=Asn runs	Merge of (ND) (N=23,812)	5	0°	20°
					5	135°	20°
Thr	Coil/NT	CH1T	X=Thr	Merge of Non(GAP)(N=134,617) and T(N=12,151)	5	0°	20°
					5	135°	20°
CysH/Cys2	Coil/NT	CH1C	X=CysH	Merge of non(GAPVINDST) (N=64,996) and C(N=2,411) (CysH and Cys2 are not distinguished)	10	0°	20°
					10	135°	20°
Gly Pre-Pro[a]	Coil, PDBr	CH2 & N2P	X=GlyPro	Merge of GlyPro(N=1,503) from Coil(withTurns) and GlyPro(N=289,407) from PDB	None		

[a]The six other (Non-Gly) Pre-Pro CMAPs were not included in the table because they were easily described in the text.

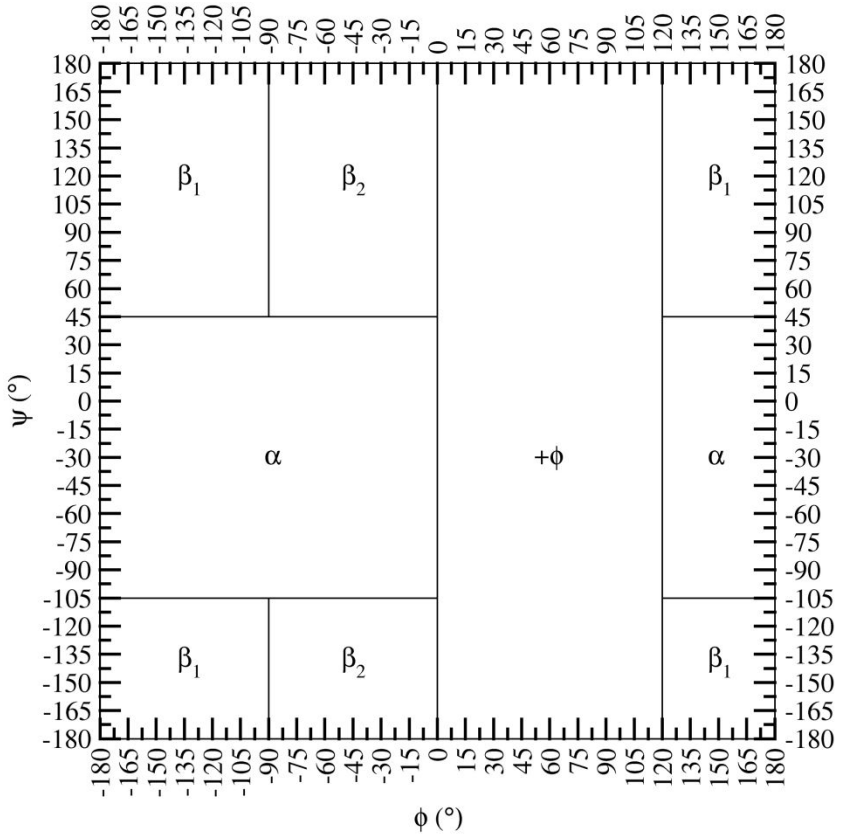


Figure S5. Definitions of Ramachandran map regions used in this work for Non(Gly). $\beta \equiv \beta_1 + \beta_2$. For Gly, $\alpha \equiv |\psi| < 75°$, $\beta \equiv |\psi| \geq 75°$.

Table S9. Definitions of Ramachandran map regions used in this work for Non(Gly)

α		β_1		β_2		+ϕ	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
$-180 \leq \phi \leq 0°$ or $\phi \geq 120°$	$-105° > \psi > 45°$	$\phi < -90°$ or $\phi > 120°$	$\psi \leq -105°$ or $\psi \geq 45°$	$-90° \leq \phi \leq 0°$	$\psi \leq -105°$ or $\psi \geq 45°$	$0° < \phi < 120°$	all ψ

$\beta \equiv \beta_1 + \beta_2$. For Gly, $\alpha \equiv |\psi| < 75°$, $\beta \equiv |\psi| \geq 75°$.

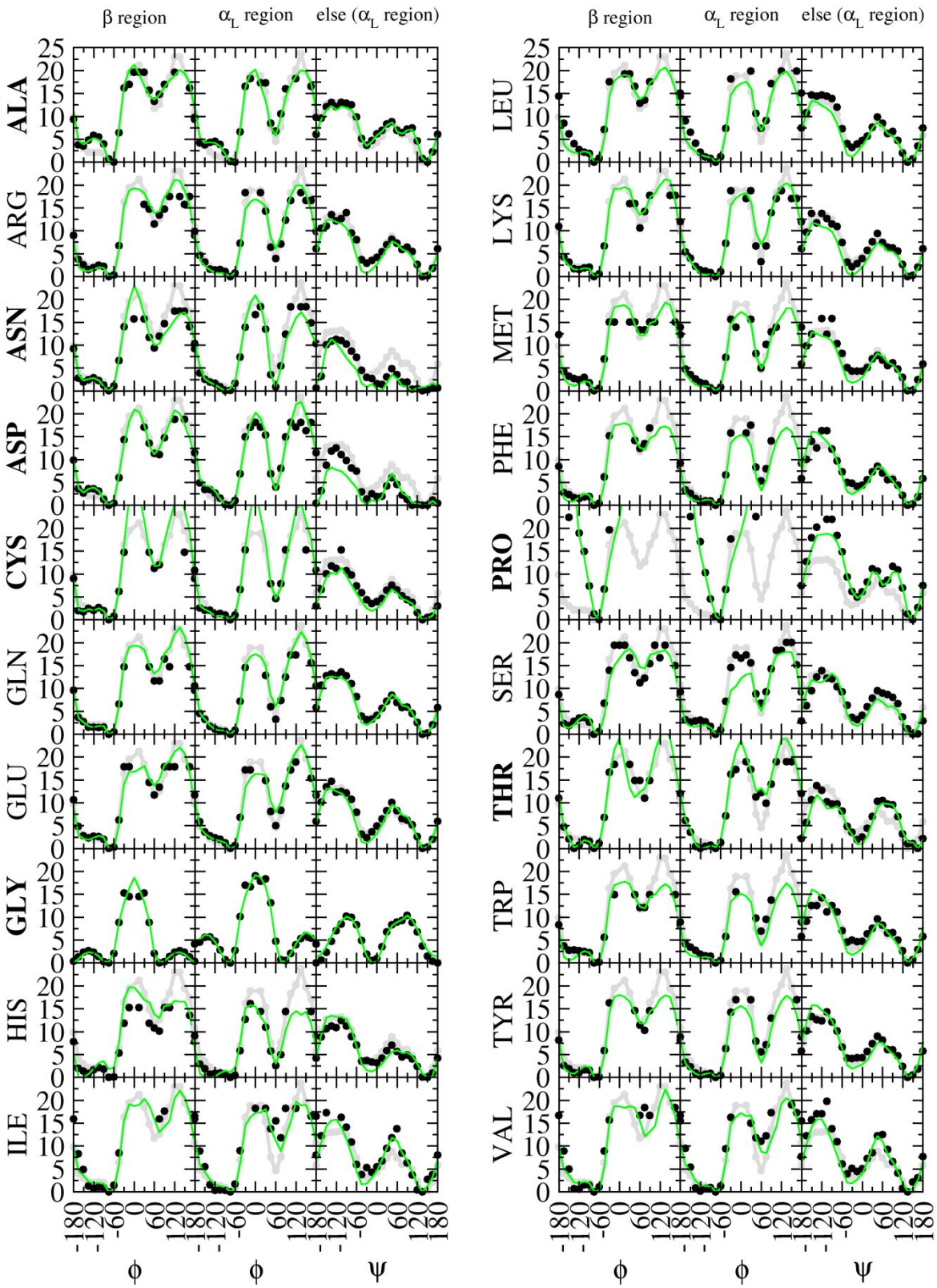


Figure S6. Restricted 1-D PMF, *W* (kJ/mol), profiles for the Coil/NT database[45] (using an effective temperature of 300 K) *vs*. KBFF20. Gray (curves with points) is the non(GAPVINDST) data from the database, black (points only) is the specific residue from the database, and green

(curves only) is the KBFF20 result from 5 μs simulations of Ac-AXA-NHM (X=all 20 amino acids). Residues labeled in bold have a specific (or nearly specific, in the case of Asn and Asp) CMAP. The regions displayed were defined as follows: "β region": ψ values ≥ 105° or ≤ -105° for all ϕ values; "α_L region": ψ values ≥ -60° across all ϕ values (so that the relative instability of α_L is apparent); "else (α_L region)": ϕ values < 0° or > 120° across all ψ values. For the creation of this figure, only one observation was required to generate a value for a given bin. For titratable residues, the simulation refers to the charged residue, and the database to a mixture. For Cys, the simulation refers to CysH, and the database to a mixture of CysH and Cys2.

Table S10. Hydration Free Energies at 298 K

Molecule	Analog For	Exp $\Delta G^*_{hydration}$ (kJ/mol)	Sim (Uncorrected) $\Delta G^*_{hydration}$ (kJ/mol)
Neopentane (NPT)[34]	N/A	10.5	10.85(5)
Isobutane (IBT)[34]	Leu/Val	9.7	9.97(6)
Propane (PRP)[34]	Val	8.2	8.34(5)
Ethane (ETH)[34]	N/A	7.6	7.77(6)
Toluene (TOL)[35]	Phe	-3.72	-5.7(3)
Methanethiol (MSH)[35]	CysH	-5.19	-6.9(1)
Dimethyl sulfide (MSM)[35]	Met	-6.74	-5.9(1)
Dimethyl disulfide (DDS)[35]	Cys2-Cys2	-7.66	-16.8(2)
Isopropanol (iPr)[35]	Thr	-19.83	-34.42(3)
Methanol (MOH)[35]	Ser	-21.34	-26.9(1)
3-Methylindole (MIN)[35]	Trp	-24.60	-33.1(2)
p-Cresol (pCr)[35]	Tyr	-25.65	-25.5(1)
Acetamide (ACT)[35]	Asn/Gln	-40.63	-41.8(1)
N-Methylacetamide (NMA)[35]	Backbone	-41.48	-39.7(2)
4-Methylimidazole (MIM)[35]	HisA/HisB	-42.97	*MIMA: -48.8(2), MIMB: -49.6(2)

*Both tautomers were simulated. MIMA is the tautomer corresponding to HisA and MIMB is the tautomer corresponding to HisB.

File S1. Example Matlab .m file (matlabInterp.m). Code created Figure S7; requires 'gridfit.m'.

```
function [] = matlabInterp()
	data = load('pmfMERGE');
	phi=data(:,1); %column 1 is phi
	psi=data(:,2); %column 2 is psi
	philin=linspace(-360,360,49);%Generates 49 points. Spacing b/w points:(720)/(48) = 15 deg.
	psilin=linspace(-360,360,49);
	[x,y]=meshgrid(philin,psilin);%Returns 2-D grid coords. based on coords. in input vectors
	figure();
	dim = [0.25 0.8 0.5 0.3]; %x_begin, y_begin, length, height
	str='Coil/NT Database, 2019, 1 observation, 15 degree bin width';
	annotation('textbox',dim,'String',str,'FitBoxToText','on','LineStyle','none',...
	'HorizontalAlignment','center','VerticalAlignment','middle');
	pmfraw=data(:,3); %column 3 has the raw Thr data
	myfunction(1,pmfraw,'Raw Thr')
	pmfraw=data(:,6); %column 6 has the raw non(GAP) data
	myfunction(3,pmfraw,'Raw non(GAP)')
	pmfraw=data(:,9); %column 9 has the raw Thr merged w/ Non(GAP) data
	myfunction(5,pmfraw,'Raw Thr Merged w/ Non(GAP)')
	print('-fillpage','Thr','-dpdf')
	function [] = myfunction(iplot,mypmf,mytitle)
		% RAW DATA
		subplot(3,2,iplot);
		zmesh=griddata(phi,psi,mypmf,x,y);
		min(mypmf)
		max(mypmf)
		myfigure(mytitle); %Make the figure of the raw data in the left column
		% SPLINED DATA
		subplot(3,2,iplot+1);
		pmfspline=mypmf;
		[zgrid,xgrid,ygrid] = gridfit(phi,psi,mypmf,philin,psilin);
		znan = isnan(mypmf);
		pmfspline(znan) = interp2(xgrid,ygrid,zgrid,phi(znan),psi(znan),'spline');
		min(pmfspline)
		max(pmfspline)
		zmesh=griddata(phi,psi,pmfspline,x,y);
		myfigure('Splined'); %Make the figure of the splined data in the right column
		function [] = myfigure(mytitle)
			zmax = 30; % kJ/mol
			zmin=  0;  % kJ/mol
			v = [0 5 10 15 20 25 30]; %kJ/mol
			xlim=[-180 180]; ylim=xlim; %degrees
			xtick=[-180 -150 -120 -90 -60 -30 0 30 60 90 120 150 180]; ytick=xtick; %deg
			zinc = 1; %kJ/mol
			zlevs = zmin:zinc:zmax;
			[C,h]=contourf(x,y,zmesh,zlevs);
			set(h,'LineColor','none'); hold on;
			zinc = 5; %kJ/mol
			zlevs = zmin:zinc:zmax;
			[C,h]=contour(x,y,zmesh,zlevs);
			set(h,'LineColor','black')
			clabel(C,h,v); hold on;
			h=colorbar;
			colormap(jet);
			ylabel(h, '{\it W} (kJ/mol)','FontName','Times New Roman','FontSize',8);
			caxis([zmin zmax]);
			set(h,'YLim',[zmin zmax],'FontSize',8,'YTick', v);
			set(gca,'XLim',xlim,'XTick', xtick, 'TickDir', 'in','FontSize',8);
			set(gca,'YLim',ylim,'YTick', ytick,'TickDir', 'in','FontSize',8);
			xlabel('{\phi}','FontName','Euclid Symbol','FontSize',12);
			xtickangle(90);
			ylabel('{\psi}','FontName','Euclid Symbol','FontSize',12);
			title(mytitle,'FontName','Times New Roman','FontSize',8);
			set(gca, 'FontName', 'Times New Roman');
			axis square;
			hold all;
		end
	end
end
```

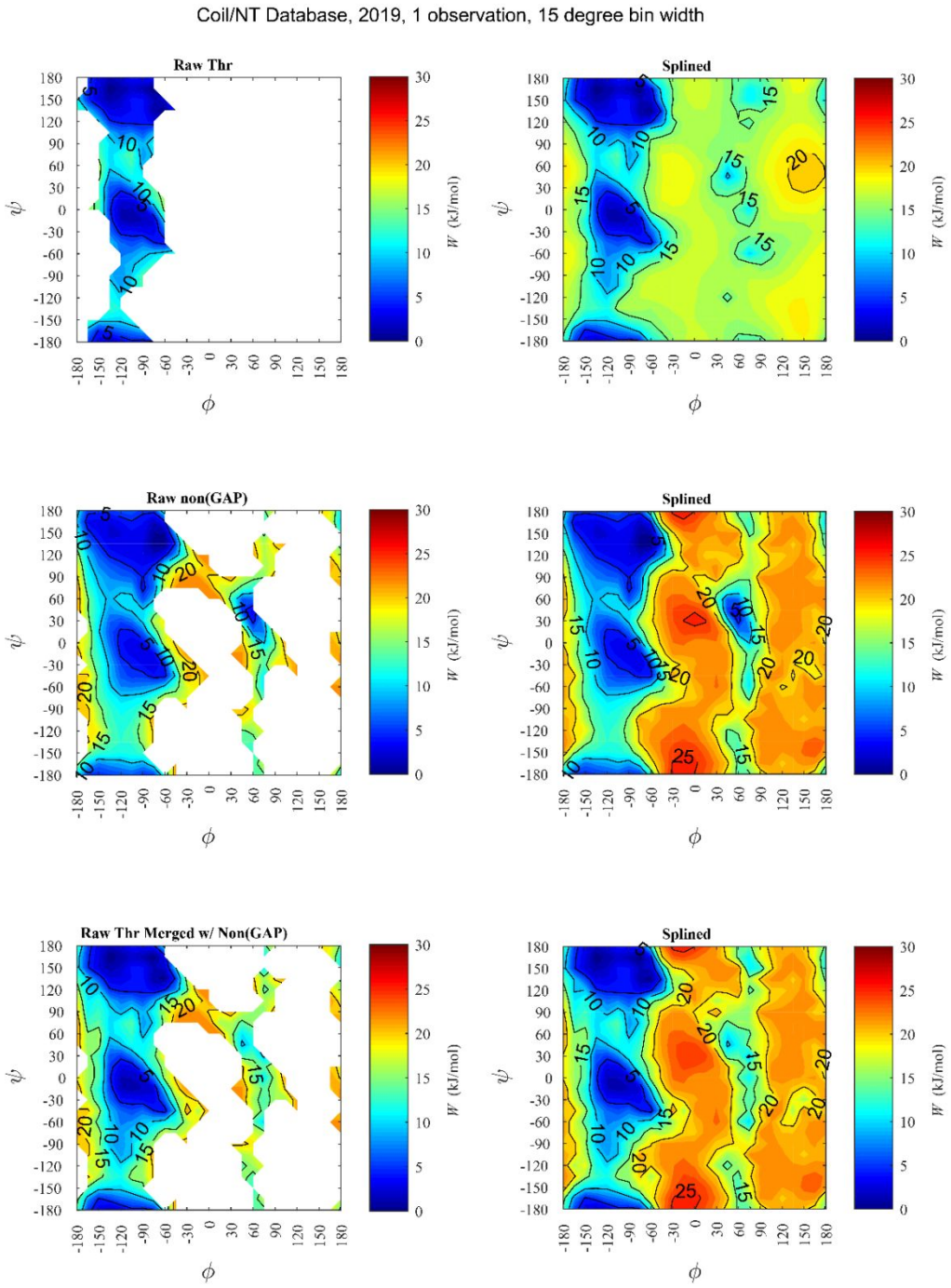


Figure S7. Example of Matlab splining using 'gridfit' to approximate barriers. Contours are of the PMF, W (kJ/mol), using $T_{\text{eff}} = 300$ K. Top, middle, bottom rows are the Thr, non(GAP), and Thr merged with non(GAP), respectively, Coil/NT database values. The left column is the raw data and the right column shows the results of the Matlab approach used to create values for the barrier regions. Had a finer resolution been used for the 'linspace' in matlabInpterp.m, contours would have been visible for the raw Thr data in the +ϕ region. The lack of barriers in the top right figure makes it clear why the Thr data could not be used on its own. Figure generated using File S1.

References

1. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. Gromacs - A Message-Passing Parallel Molecular-Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43-56.
2. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
3. Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B. *GROMACS User Manual version 2016*, 2018.
4. Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255-268.
5. Hoover, W. G. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
6. Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604-1608.
7. Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. Replica-Exchange Monte Carlo Method for the Isobaric-Isothermal Ensemble. *Chem. Phys. Lett.* **2001**, *335*, 435-439.
8. Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. Reproducible Polypeptide Folding and Structure Prediction Using Molecular Dynamics Simulations. *J. Mol. Biol.* **2005**, *354*, 173-183.
9. van der Spoel, D.; Seibert, M. M. Protein Folding Kinetics and Thermodynamics from Atomistic Simulations. *Phys. Rev. Lett.* **2006**, *96*, 238102.
10. Lei, H. X.; Duan, Y. Improved Sampling Methods for Molecular Simulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 187-191.
11. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - A New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.
12. Nose, S.; Klein, M. L. Constant Pressure Molecular-Dynamics for Molecular-Systems. *Mol. Phys.* **1983**, *50*, 1055-1076.
13. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463-1472.
14. Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952-962.
15. Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High-Resolution Computer Models of a Plasma. *J. Comput. Phys.* **1974**, *14*, 148-158.
16. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
17. Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties. *J. Chem. Theory Comput.* **2013**, *9*, 3527-3537.
18. Wennberg, C. L.; Murtola, T.; Pall, S.; Abraham, M. J.; Hess, B.; Lindahl, E. Direct-Space Corrections Enable Fast and Accurate Lorentz-Berthelot Combination Rule Lennard-Jones Lattice Summation. *J. Chem. Theory Comput.* **2015**, *11*, 5737-5746.
19. Páll, S.; Hess, B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Comput. Phys. Commun.* **2013**, *184*, 2641-2650.
20. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.
21. Widom, B. Some Topics in Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808-2812.

22. Hess, B. Determining the Shear Viscosity of Model Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 209-217.
23. Smith, P. E.; Van Gunsteren, W. F. The Viscosity of SPC and SPC/E Water at 277-K and 300-K. *Chem. Phys. Lett.* **1993**, *215*, 315-318.
24. Williams, W. D.; Ellard, J. A.; Dawson, L. R. Solvents Having High Dielectric Constants .6. Diffusion in N-Methylacetamide. *J. Am. Chem. Soc.* **1957**, *79*, 4652-4654.
25. Celebi, A. T.; Jamali, S. H.; Bardow, A.; Vlugt, T. J. H.; Moultos, O. A. Finite-Size Effects of Diffusion Coefficients Computed From Molecular Dynamics: A Review of What We Have Learned So Far. *Mol. Simul.* **2020**, 1-15.
26. Kang, M.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Amides. *J. Comput. Chem.* **2006**, *27*, 1477-1485.
27. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Sodium Chloride in Water. *J. Chem. Phys.* **2003**, *119*, 11342-11349.
28. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Methanol and Aqueous Methanol Solutions. *J. Phys. Chem. B* **2005**, *109*, 15080-15086.
29. Ploetz, E. A.; Rustenburg, A. S.; Geerke, D. P.; Smith, P. E. To Polarize or Not to Polarize? Charge-on-Spring versus KBFF Models for Water and Methanol Bulk and Vapor-Liquid Interfacial Mixtures. *J. Chem. Theory Comput.* **2016**, *12*, 2373-2387.
30. D'Errico, J. *Surface Fitting Using Gridfit (https://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit), MATLAB Central File Exchange.*, 1.1.0.0 (Updated 04 Mar 2016); 2020.
31. Patriksson, A.; van der Spoel, D. A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073-2077.
32. Patriksson, A.; van der Spoel, D. Temperature Generator for REMD-Simulations. http://folding.bmc.uu.se/remd/ (accessed 9/1/2020).
33. Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
34. Graziano, G. On the Size Dependence of Hydrophobic Hydration. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 3345-3352.
35. Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. Optimizing Solute-Water van der Waals Interactions To Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012**, *116*, 4524-4534.
36. Shapovalov, M. V.; Dunbrack, R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*, 844-858.
37. Shapovalov, M. V.; Dunbrack, R. L., Dataset Used in Computing of '2010 Backbone-Dependent Rotamer Library'. Fox Chase Cancer Center, Philadelphia, PA, USA, 2012.
38. Lovell, S. C.; Davis, I. W.; Adrendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by C Alpha Geometry: Phi,Psi and C Beta Deviation. *Proteins* **2003**, *50*, 437-450.
39. Suh, S.; Ploetz, E. A.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Zwitterionic Amino Acids. *To be submitted*.
40. Ploetz, E. A.; Weerasinghe, S.; Kang, M.; Smith, P. E. Accurate Force Fields for Molecular Simulation. In *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics*, 1st ed.; Smith, P. E.; Matteoli, E.; O'Connell, J. P., Eds. CRC Press: Boca Raton, 2013; pp 117-131.

41. Gee, M. B.; Cox, N. R.; Jiao, Y. F.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Aqueous Alkali Halides. *J. Chem. Theory Comput.* **2011**, *7*, 1369-1380.
42. Naleem, N.; Bentenitis, N.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Alkaline Earth Halide Salts. *J. Chem. Phys.* **2018**, *148*, 222828.
43. Wang, G. L.; Dunbrack, R. L. PISCES: A Protein Sequence Culling Server. *Bioinformatics* **2003**, *19*, 1589-1591.
44. Fitzkee, N. C.; Fleming, P. J.; Rose, G. D. The Protein Coil Library: A Structural Database of Nonhelix, Nonstrand Fragments Derived from the PDB. *Proteins* **2005**, *58*, 852-854.
45. Fitzkee, N. C.; Fleming, P. J.; Rose, G. D., pisces_20190513_pc20_res1.6_R0.25. 2019.
46. Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980-980.
47. Hughes, E. W.; Lipscomb, W. N. The Crystal Structure of Methylammonium Chloride. *J. Am. Chem. Soc.* **1946**, *68*, 1970-1975.
48. Wagner, W.; Pruss, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387-535.
49. Harvey, A. H.; Peskin, A. P.; Klein, S. A. *NIST Standard Reference Database 10: NIST/ASME Steam Properties*, Version 2.22; U.S. Department of Commerce: Gaithersburg, 2008.
50. Hsu, L. Y.; Nordman, C. E. Structures of 2 Forms of Sodium-Acetate, Na+.C2H3O2-. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1983**, *39*, 690-694.
51. Zakharov, B. A.; Tumanov, N. A.; Boldyreva, E. V. Beta-Alanine Under Pressure: Towards Understanding the Nature of Phase Transitions. *CrystEngComm* **2015**, *17*, 2074-2079.
52. Moggach, S. A.; Allan, D. R.; Parsons, S.; Sawyer, L. Effect of Pressure on the Crystal Structure of Alpha-Glycylglycine to 4.7 GPa; Application of Hirshfeld Surfaces to Analyse Contacts on Increasing Pressure. *Acta Crystallogr., Sect. B: Struct. Sci.* **2006**, *62*, 310-320.
53. Haas, D. J.; Harris, D. R.; Mills, H. H. Crystal Structure of Guanidinium Chloride. *Acta. Crystallogr.* **1965**, *19*, 676-679.
54. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for the Simulation of Aqueous Guanidinium Chloride Solutions. *J. Chem. Phys.* **2004**, *121*, 2180-2186.
55. Krumbe, W.; Haussuhl, S. Crystal-Culture and Determination of Structure for Guanidinium-Hydrogen-Selenite, Guanidinium-Hydrogen-Phosphite, Guanidinium-Tetrafluoroborate, Guanidinium-Glutarate and Guanidinium-Acetate. *Z. Kristallogr.* **1987**, *178*, 132-134.
56. Vadivel, S.; Sultan, A. B.; Samad, S. A.; Shunmuganarayanan, A.; Muthu, R. Synthesis, Structural Elucidation, Thermal, Mechanical, Linear and Nonlinear Optical Properties of Hydrogen Bonded Organic Single Crystal Guanidinium Propionate for Optoelectronic Device Application. *Chem. Phys. Lett.* **2018**, *707*, 165-171.
57. Koo, C. H.; Lee, O. J.; Sin, H. S. The Crystal Structure of Monoethanolamine Hydrochloride. *J. Korean Chem. Soc.* **1972**, *16*, 6-12.
58. Penttila, A.; Uusi-Kyyny, P.; Salminen, A.; Seppala, J.; Alopaeus, V. A Comprehensive Thermodynamic Study of Heat Stable Acetic Acid Salt of Monoethanolamine. *Int. J. Greenhouse Gas Control* **2014**, *22*, 313-324.

59. Hosseini, S. M.; Hosseinian, A.; Aparicio, S. An Experimental and Theoretical Study on 2-Hydroxyethylammonium Acetate Ionic Liquid. *J. Mol. Liq.* **2019**, *284*, 271-281.
60. Lemmon, E. W.; McLinden, M. O.; Friend, D. G., Thermophysical Properties of Fluid Systems. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, Linstrom, P. J.; Mallard, W. G., Eds. National Institue of Standards and Technology: Gaithersburg, MD 20899, (retrieved July 24, 2020).
61. Zielkiewicz, J. Excess Volumes in (N-Methylacetamide Plus Methanol Plus Water) at the Temperature 313.15 K. *J. Chem. Thermodyn.* **1999**, *31*, 1597-1604.
62. Diazpena, M.; Tardajos, G. Isothermal Compressibilities of Normal-1-Alcohols from Methanol to 1-Dodecanol at 298.15-K, 308.15-K, 318.15-K, and 333.15-K. *J. Chem. Thermodyn.* **1979**, *11*, 441-445.
63. Boodida, S.; Bachu, R. K.; Patwari, M. K.; Nallani, S. Volumetric and Transport Properties of Binary Liquid Mixtures of N-Methylacetamide with Lactones at Temperatures (303.15 to 318.15) K. *J. Chem. Thermodyn.* **2008**, *40*, 1422-1427.
64. Chaudhary, N.; Nain, A. K. Densities, Ultrasonic Speeds, Viscosities, Refractive Indices, and Excess Properties of 1-Butyl-3-methylimidazolium Tetrafluoroborate plus N-Methylacetamide Binary Mixtures at Different Temperatures. *J. Chem. Eng. Data* **2020**, *65*, 1447-1459.
65. Hatibarua, J.; Parry, G. S. Crystallographic Study of Acetates of Potassium, Rubidium and Cesium. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1972**, *B 28*, 3099-3100.
66. Verrall, R. E.; Conway, B. E. Partial Molar Volumes and Adiabatic Compressibilities of Tetraalkylammonium and Aminium Salts in Water .2. Volume and Volume Change Relationships. *J. Phys. Chem.* **1966**, *70*, 3961-3969.
67. Bonner, O. D. Osmotic and Activity-Coefficients of Methyl-Substituted Ammonium Chlorides. *J. Chem. Soc., Faraday Trans. 1* **1981**, *77*, 2515-2518.
68. Novotny, P.; Sohnel, O. Densities of Binary Aqueous-Solutions of 306 Inorganic Substances. *J. Chem. Eng. Data* **1988**, *33*, 49-55.
69. Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*. 2nd ed.; Butterworths Publications Limited: London, 1959; p 559.
70. Gucker, F. T.; Allen, T. W. The Densities and Specific Heats of Aqueous Solutions of dl-α-Alanine, β-Alanine and Lactamide1,2. *J. Am. Chem. Soc.* **1942**, *64*, 191-199.
71. Smith, E. R. B.; Smith, P. K. Thermodynamic Properties of Solutions of Amino Acids and Related Substances: IV. The Effect of Increasing Dipolar Distance on the Activities of Aliphatic Amino Acids in Aqueous Solution at Twenty-Five Degrees. *J. Biol. Chem.* **1940**, *132*, 47-56.
72. Dyke, S. H.; Hedwig, G. R.; Watson, I. D. Relative Partial Molar Enthalpies and Apparent Molar Volumes of Dipeptides in Aqueous-Solution. *J. Solution Chem.* **1981**, *10*, 321-331.
73. Smith, E. R. B.; Smith, P. K. Thermodynamic Properties of Solutions of Amino Acids and Related Substances: VI. The Activities of Some Peptides in Aqueous Solution at Twenty-Five Degrees. *J. Biol. Chem.* **1940**, *135*, 273-279.
74. Makhatadze, G. I.; Fernandez, J.; Freire, E.; Lilley, T. H.; Privalov, P. L. Thermodynamics of Aqueous Guanidinium Hydrochloride Solutions in the Temperature-Range from 283.15 to 313.15-K. *J. Chem. Eng. Data* **1993**, *38*, 83-87.
75. Pitzer, K. S. Thermodynamics of Electrolytes .1. Theoretical Basis and General Equations. *J. Phys. Chem.* **1973**, *77*, 268-277.

76. Miyajima, K.; Inari, K.; Hamaguchi, N.; Yoshida, H.; Nakagaki, M. Studies on Aqueous-Solutions of Guanidinium Salts .2. Activity-Coefficients and Partial Molar Volumes for Guanidinium Salts in Aqueous-Solutions at 25-Degrees C. *Nippon Kagaku Kaishi* **1975**, 1447-1452.
77. Miyajima, K.; Yoshida, H.; Nakagaki, M. Studies on Aqueous-Solutions of Guanidinium Salts .7. Activity-Coefficients and Partial Molar Volumes for Guanidinium Salts with Hydrophobic Counter Anions in Aqueous-Solutions. *Bull. Chem. Soc. Jpn.* **1978**, *51*, 2508-2512.
78. Lee, L. S.; Lee, C. C. Vapor Pressures and Enthalpies of Vaporization of Aqueous Solutions of Triethylammonium Chloride, 2-Hydroxyethylammonium Chloride, and Tris(hydroxymethyl)aminomethane Hydrochloride. *J. Chem. Eng. Data* **1998**, *43*, 469-472.
79. McCullough, J. P.; Pennington, R. E.; Waddington, G. A Calorimetric Determination of the Vapor Heat Capacity and Gas Imperfection of Water. *J. Am. Chem. Soc.* **1952**, *74*, 4439-4442.
80. Nasirzadeh, K.; Neueder, R. Measurements and Correlation of Osmotic Coefficients and Evaluation of Vapor Pressure for Solutions of KCH3COO and NACH(3)COO in Methanol at 25 Degrees C. *Acta Chim. Slov.* **2004**, *51*, 117-126.
81. Zielkiewicz, J. (Vapour Plus Liquid) Equilibrium Measurements and Correlation of the Ternary Mixture (N-Methylacetamide Plus Methanol Plus Water) at the Temperature 313.15 K. *J. Chem. Thermodyn.* **1999**, *31*, 819-825.
82. Kreis, R. W.; Wood, R. H. Freezing Points, Osmotic Coefficients, and Activity Coefficients of Salts in N-Methylacetamide .2. Tetraalkylammonium Halides and Some Alkali Metal Formates, Acetates, and Propionates. *J. Phys. Chem.* **1971**, *75*, 2319-2325.
83. Zielkiewicz, J. Preferential Solvation of N-Methylformamide, N,N-Dimethylformamide and N-Methylacetamide by Water and Alcohols in the Binary and Ternary Mixtures. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2925-2932.
84. Gregoire, F.; Wei, S. H.; Streed, E. W.; Brameld, K. A.; Fort, D.; Hanely, L. J.; Walls, J. D.; Goddard, W. A.; Roberts, J. D. Conformational Equilibria of Beta-Alanine and Related Compounds as Studied by NMR Spectroscopy. *J. Am. Chem. Soc.* **1998**, *120*, 7537-7543.
85. Smith, T. D.; Gerken, J. B.; Jog, P. V.; Roberts, J. D. Conformational Equilibria of Ethanolamine and its Hydrochloride in Solution. *Org. Lett.* **2007**, *9*, 4555-4557.
86. Armstrong, D. A.; Kaas, Q.; Rosengren, K. J. Prediction of Disulfide Dihedral Angles Using Chemical Shifts. *Chem. Sci.* **2018**, *9*, 6548-6556.
87. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
88. Terhorst, J. P.; Jorgensen, W. L. E/Z Energetics for Molecular Modeling and Design. *J. Chem. Theory Comput.* **2010**, *6*, 2762-2769.