Supporting Information for The ADD Force Field for Sugars and Polyols: Predicting the Additivity of Protein-Osmolyte Interaction

Andrea Arsiccio,[†] Pritam Ganguly,[†] Lorenzo La Cortiglia,[‡] Joan-Emma Shea,[†,¶] and Roberto Pisano[*,‡]

[†]*Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States*

[‡]*Department of Applied Science and Technology, Politecnico di Torino, 24 corso Duca degli Abruzzi, Torino 10129, Italy*

[¶]*Department of Physics, University of California, Santa Barbara, California 93106, United States*

E-mail: roberto.pisano@polito.it

Partial Charges of Simulated Carbohydrates

In the KBP[1] force field the partial charges of alcohol O and H are -0.50 and 0.18, respectively. The magnitude of these charges is increased in the ADD parameterization to -0.65 and 0.33, respectively. The non-alcohol atoms partial charges for the KBP and ADD force fields are the same, and are shown in Figure S1 for the two molecules (sucrose and sorbitol) considered in this work.

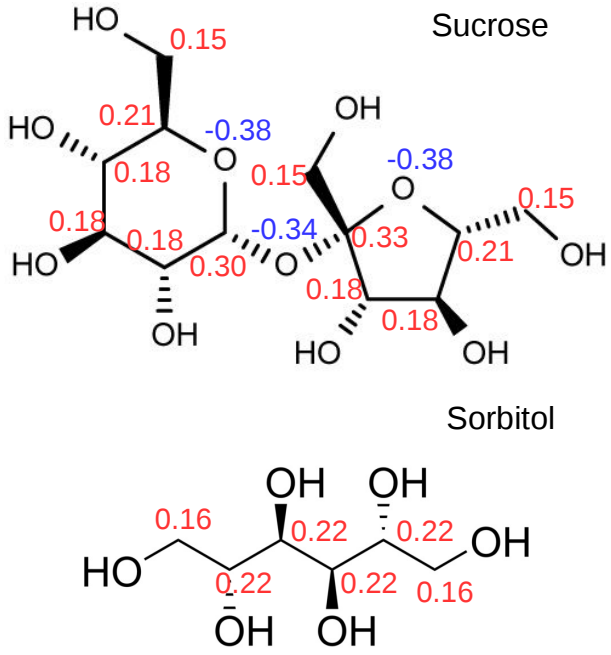

Figure S1: Partial charges of sucrose and sorbitol carbon (red) and non-alcohol oxygen (blue) atoms in the KBP and ADD force fields. The partial charge for all alkyl hydrogens is 0.09.

Sidechain Contributions: Comparison between Zwitterionic and Capped Amino Acids

Following the approach described by Auton et al.,[2,3] it is assumed that additivity exists also within the single amino acid. The sidechain contribution γ^{sc} can therefore be calculated by subtracting the KB integral γ_{GLY} for glycine (where the sidechain is just a hydrogen atom)

to the KB integral of the amino acid i being considered γ_i,

$$\gamma_i^{sc} = \gamma_i - \gamma_{GLY} \tag{1}$$

To verify that the sidechain contribution would not vary depending on the terminal capping conditions, both capped (acetylated and amidated) and zwitterionic amino acids were considered (for both sucrose and sorbitol, sim. 1 and 2 in Table 1 of the main text). The resulting γ^{sc} values are shown in Figure S2a-c, for the KBP and ADD force fields.

It was observed that the trend was essentially the same for both capped and zwitterionic amino acids. A change in sign was noticed only for the case of serine and the KBP force field, but it should be considered that the γ^{sc} value in this case is close to zero for both the zwitterionic and the capped form, and the absolute difference between the two values is not significant. The simulation results, therefore, seem to confirm the existence of additivity within the individual amino acid (i.e., amino acid KB integral = backbone contribution + side chain contribution). This also means that, in our specific case, both zwitterionic and capped amino acids may be considered to compute the side chain contribution, without dramatically affecting the final results. In the main text, attention was focused on zwitterionic residues, in line with what was done experimentally.[4]

The KBP Parameters Lead to Demixing of Sucrose-Guanidinium Chloride Mixtures

Guanidinium chloride (GdmCl) is a strong denaturant, commonly used in studies of protein folding. For instance, the m value of protecting osmolytes, such as sucrose, is often extracted from its ability to counteract chemical denaturation of proteins induced by GdmCl. The guanidinium group is also encountered in the side chain of arginine, and is responsible for the polar behavior of this amino acid.

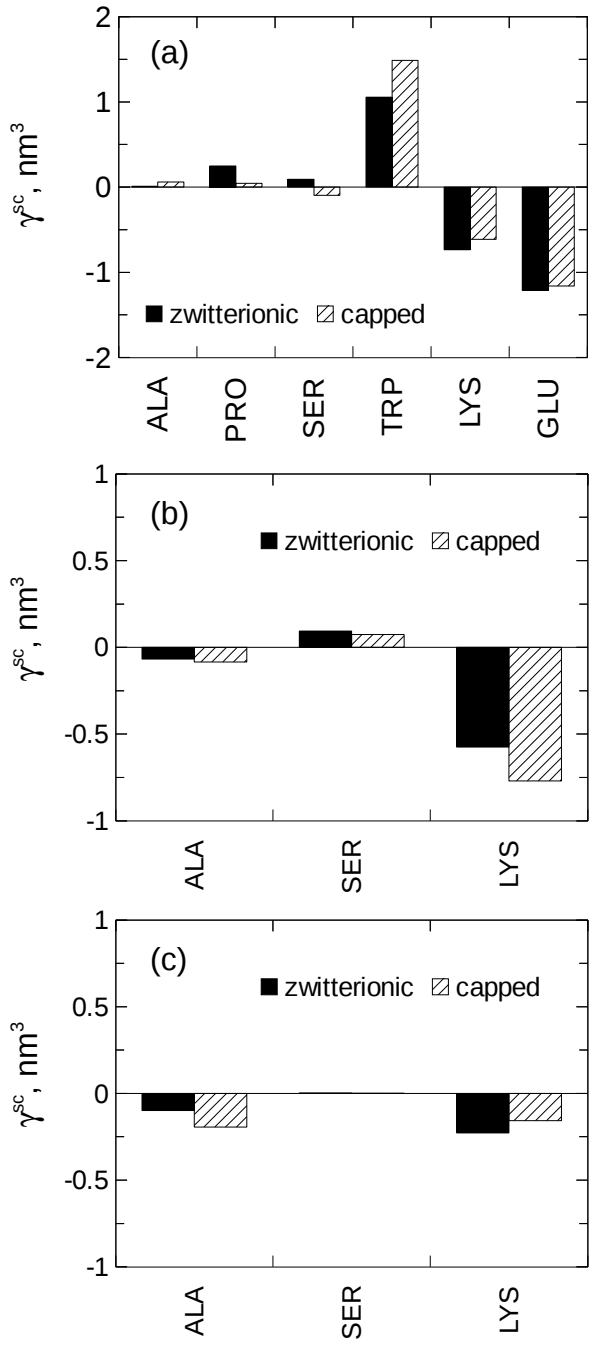

Figure S2: Sidechain KB integrals ($\gamma_i^{sc} = \gamma_i - \gamma_{GLY}$) for both zwitterionic (black bars) and capped (dashed bars) amino acids in (a) 1 M sucrose for the KBP force field, (b) 1 M sucrose for the ADD force field, (c) 1 M sorbitol for the ADD force field.

Simulations of 1 M sucrose in presence of 4 M GdmCl were performed. Sucrose was described with either the KBP[1] or ADD force field, while a modified version of the Smith force field,[5] derived by Wernersson et al.,[6] was used for GdmCl. This GdmCl force field was originally developed using the geometric mean rule for the van der Waals cross-interactions. We performed the simulations using both the Lorentz-Berthelot (LB) rule (which is the default for CHARMM36) and the geometric mean (G) rule. In the case of the LB combination rule, Lennard-Jones parameters of a cross-interaction are defined as the arithmetic (for σ) or geometric (for ε) mean of the individual atomic values. The geometric combination rule instead prescribes the use of a geometric mean for both σ and ε.

A cubic simulation box with 6 nm side length, containing 130 sucrose molecules, 520 GdmCl ions and 2843 water molecules was built. Water was described using the CHARMM TIP3P water model.[7] The simulations were run for 60 ns at 1 bar and 300 K, and the last 40 ns were used for the analysis of Kirkwood-Buff integrals (KBIs). Gdm^+ and Cl^- were treated as indistinguishable ions during the calculations. All the other simulation details were the same described in the Materials and Methods section of the main text.

The results obtained for these simulations are shown in Table S1, where the following notation was used: i = ions, w = water, s = sucrose.

Table S1: Comparison between the ion-ion (G_{ii}), ion-sucrose (G_{is}), ion-water (G_{iw}), sucrose-sucrose (G_{ss}) and sucrose-water (G_{sw}) KBIs as obtained experimentally and as predicted by the ADD or KBP force fields, with both the geometric (G) and Lorentz-Berthelot (LB) combination rules.

4 M GdmCl + 1 M Sucrose				
KBIs	**KBP-G**	**KBP-LB**	**ADD-G**	**ADD-LB**
G_{ii}, nm^3	0.683	0.237	0.113	-0.058
G_{is}, nm^3	-1.589	-0.945	-0.406	-0.228
G_{iw}, nm^3	0.292	0.227	0.003	0.031
G_{ss}, nm^3	1.824	0.860	-0.582	-0.824
G_{sw}, nm^3	-0.784	-0.659	-0.149	-0.181

From Table S1 it is clear that both KBP and ADD sucrose are excluded from GdmCl

(negative G_{is}). This agrees with the fact that both force fields predict exclusion of sucrose from charged amino acids (see Figure 2 in the main text). However, in presence of GdmCl, KBP sucrose tends to self-interact (positive G_{ss}), and the very unfavorable sucrose-ion interaction promotes the ion-ion interactions (positive G_{ii}). This combination of KBIs may lead to demixing of the solution, and such a phenomen was observed, for instance, during the KBP-G simulation (Figure S3). This demixing may lead to artefact during simulations, and suggests incompatibility between the KBP and GdmCl force fields.

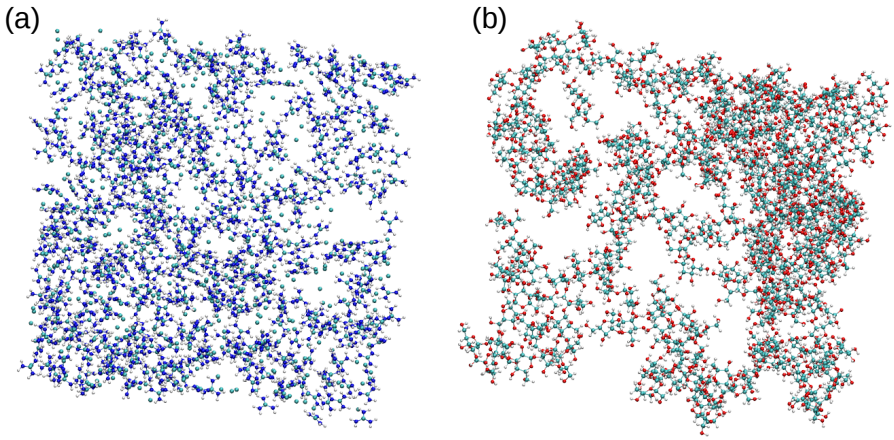

Figure S3: Snapshots of the 4 M GdmCl + 1 M sucrose simulation box obtained for the KBP force field, and with the geometric combination rule. The GdmCl ions cluster on the left-hand side of the box (panel a) while sucrose molecules are confined to the right-hand side (panel b). A demixing of the solution is therefore observed.

The value of G_{is} is less negative when ADD sucrose is used. This is also in accordance with the lower exclusion of sucrose from the side chain of arginine that was observed using the ADD parameters instead of the KBPs (see Figure 2 in the main text). This translates into less positive (ADD-G) or even negative (ADD-LB) values of G_{ii}. Moreover, the sucrose self-interaction (G_{ss}) is always unfavorable when using the ADD force field. The negative values of G_{ii} and G_{ss} for the ADD-LB combination suggest that demixing should be avoided in these conditions. In line with this, no demixing was observed during the ADD-LB simulation.

Overall, the use of the geometric rule instead of the Lorentz-Berthelot rule makes the ion-ion and sucrose-sucrose interactions more favorable, penalizing the sucrose-ion cross-

interactions.

ADD and KBP Behavior in the Case of Different Combination Rules

The CHARMM force field uses the Lorentz-Berthelot combination rule for the van der Waals cross-interactions. Simulations 4 in Table 1 of the main text were also performed using the geometric combination rule (i.e., geometric mean for both σ and ε), for the case of sucrose as a model osmolyte. A cubic box with 6 nm side length and containing 130 sucrose molecules was built. The solution KBIs (G_{33} and G_{13}) were computed, using the last 40 ns of the 60-ns long trajectories. All the other simulation details were the same described in the Materials and Methods section of the main text. The results for both combination rules and for both KBP and ADD are shown in Table S2.

Table S2: Comparison between the sugar-sugar (G_{33}), and sugar-water (G_{13}) KBIs as obtained experimentally,[8,9] or as predicted by the ADD and KBP force fields with the geometric (G) and Lorentz-Berthelot (LB) rules.

1 M Sucrose					
KBIs	**KBP-G**	**KBP-LB**	**ADD-G**	**ADD-LB**	**Exp.**
G_{33}, nm^3	-0.436	-0.843	-0.766	-0.758	-0.819
G_{13}, nm^3	-0.360	-0.251	-0.201	-0.193	-0.220

The behavior of ADD sucrose was not significantly modified when changing the combination rule for the van der Waals cross-interactions. The results obtained with the ADD force field always remained close to the experimental values. In contrast, the KBP sucrose self-interaction became less unfavorable when the geometric mean was used instead of the Lorentz-Berthelot rule, leading to a significant deviation from experiments.

Diffusion Coefficients of Sucrose and Sorbitol

The diffusion coefficients of sucrose and sorbitol, as predicted by the original CHARMM36, KBP or ADD force fields, were also computed. For this purpose, three different concentrations were investigated for both sucrose (0.5 M, 1 M, and 1.4 M) and sorbitol (0.1 M, 0.5 M, and 1 M), and the experimental data reported in[10] and[11] were used for reference.

For these simulations, cubic boxes with 4 nm side length were built, filled in with the appropriate number of carbohydrate molecules to give the desired concentration, and then solvated with CHARMM TIP3P water.[7] All the remaining computational details, including equilibration and production procedure, thermostats/barostats used and simulation times were the same already described in the Materials and Methods section of the main text.

The diffusion coefficients were then computed as described in Hatcher et al.[12] D_{PBC} was calculated by fitting a straight line through the mean square displacement of the carbohydrate molecules as function of time. The obtained diffusivity was then corrected to account for the underestimation of water viscosity by the CHARMM TIP3P model used,

$$D = \left(D_{PBC} + \frac{k_B T \zeta}{6\pi\eta L} \right) \cdot \frac{\eta_{\text{CHARMM TIP3P}}}{\eta_{exp}} \tag{2}$$

where k_B is the Boltzmann constant, T is the absolute temperature (300 K in our case), ζ is a constant of 2.837297, η is the viscosity and L is the length of the cubic simulation box. In particular, η was calculated from the Einstein formula,

$$\eta = \eta_{\text{CHARMM TIP3P}}(1 + 2.5\phi) \tag{3}$$

where $\eta_{\text{CHARMM TIP3P}}$ is the viscosity of CHARMM TIP3P water at 300 K, and ϕ is the volume fraction of the solute within the box.

The value of $\eta_{\text{CHARMM TIP3P}}$ was here determined from a preliminary simulation. For this purpose, a cubic box, with 8 nm side length, was filled in with CHARMM TIP3P water molecules, energy minimzed with the steepest descent algorithm, and equilibrated for 1 ns

at 300 K and 1 bar using Berendsen temperature (1 ps relaxation time) and pressure (1 ps relaxation time) coupling.[13] The system was subsequently simulated for 500 ps at 300 K and 1 bar, controlling temperature and pressure with the Nosé-Hoover thermostat[14,15] (0.5 ps relaxation time) and Parrinello-Rahman barostat[16] (3 ps relaxation time), respectively. Periodic boundary conditions were used, the water molecules were kept rigid using the SETTLE algorithm,[17] the cut-off radius for both Coulombic and Lennard-Jones interactions was set to 1.2 nm. During the 500-ps run, a cosinusoidal acceleration, with 0.02 nm ps^{-2} amplitude, was applied to the box to compute viscosity, according to the periodic perturbation method.[18] A value of $\eta_{\text{CHARMM TIP3P}} = 0.32$ cP was obtained from this run, and used in Equation 3. The value of $\eta_{\text{CHARMM TIP3P}}$ was also divided by the experimental viscosity of water at 300 K (0.85 cP) to yield the correction term $\eta_{\text{CHARMM TIP3P}}/\eta_{exp} = 0.376$ of Equation 2. The results of this analysis are shown in Figure S4.

As already observed in[1] we found that the KBP parameters result in higher diffusion coefficients compared to the original CHARMM36 force field. ADD carbohydrates generally show intermediate values, albeit closer to those observed for the KBPs. For sorbitol (Figure S4b), the ADD and KBP parameters improve the description of the experimental trend, while the opposite is true for sucrose (Figure S4a).

Compatibility of ADD with the AMBER 99SB-ILDN and OPLS-AA Force Fields

Trpzip1 (SWTWEGNKWTWK) (pdb 1LE0[19]) was simulated in presence of 1 M sucrose, in a 6.1 nm cubic box. The peptide was capped by an acetyl group and an amide moiety at the N and C termini, respectively. Sucrose molecules were described with the ADD force field, while the CHARMM36m,[20] AMBER 99SB-ILDN,[21] or OPLS-AA[22] force fields were used for the protein. CHARMM TIP3P[7] was used in combination with CHARMM36m, while the original TIP3P model[23] was used with AMBER and OPLS. Simulations were performed at

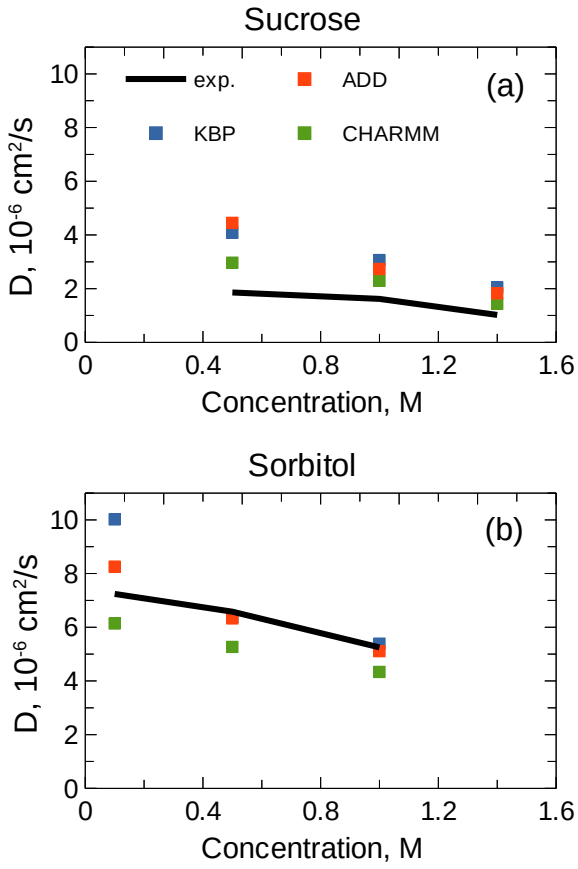

Figure S4: Diffusion coefficients of (a) sucrose and (b) sorbitol as function of concentration, as measured experimentally or computed with the orginal CHARMM36, KBP and ADD force fields.

pH 7 (+1 charge, neutralized by the addition of Cl^- ions), 300 K and 1 bar, and run for 150 ns. Details for the equilibration and production run are the same described in the Materials and Methods section of the main text. The last 100 ns were used for analysis.

The preferential interaction coefficient $\Gamma(r)$ (Eq. 12 in the main text) was computed, and is shown in Figure S5 as function of the distance from the protein surface. Considering that the measurement of preferential solvation is characterized by some uncertainty, the results we obtained for the different protein force fields are in the range of the deviation which is generally expected for this type of analysis. This suggests that the ADD parameters behave similarly when CHARMM, AMBER or OPLS are used, showing their compatibility with these widely used force fields.

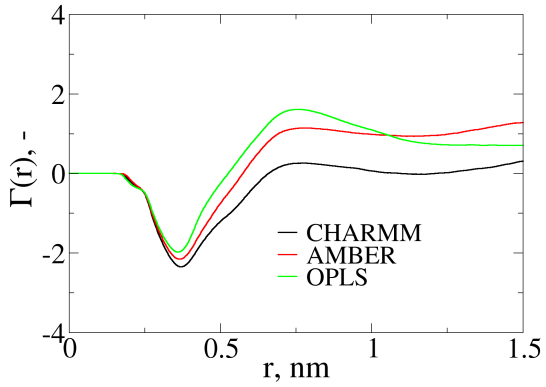

Figure S5: Running value of the preferential interaction coefficient $\Gamma(r)$, computed for Trpzip1 in presence of ADD sucrose, at 1 M concentration. The CHARMM36m, AMBER 99SB-ILDN and OPLS-AA force fields were used for the peptide.

References

(1) Cloutier, T.; Sudrik, C.; Sathish, H. A.; Trout, B. L. Kirkwood–Buff-derived alcohol parameters for aqueous carbohydrates and their application to preferential interaction coefficient calculations of proteins. *J. Phys. Chem. B* **2018**, *122*, 9350–9360.

(2) Auton, M.; Bolen, D. W. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15065–15068.

(3) Auton, M.; Bolen, D. W.; Rösgen, J. Structural thermodynamics of protein preferential solvation: Osmolyte solvation of proteins, aminoacids, and peptides. *Proteins* **2008**, *73*, 802–813.

(4) Liu, Y.; Bolen, D. W. The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry* **1995**, *34*, 12884–12891.

(5) 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.

(6) Wernersson, E.; Heyda, J.; Vazdar, M.; Lund, M.; Mason, P. E.; Jungwirth, P. Orientational Dependence of the Affinity of Guanidinium Ions to the Water Surface. *J. Phys. Chem. B* **2011**, *115*, 12521–12526.

(7) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(8) Sangster, J.; Teng, T.-T.; Lenzi, F. Molal volumes of sucrose in aqueous solutions of NaCl, KCl, or urea at 25°C. *J. Solution Chem. 5*, 575–585.

(9) Robinson, R. A.; Stokes, R. H. Activity coefficients in aqueous solutions of sucrose, mannitol and their mixtures at 25°C. *J. Phys. Chem.* **1961**, *65*, 1954–1958.

(10) Ekdawi-Sever, N.; de Pablo, J. J.; Feick, E.; von Meerwall, E. Diffusion of Sucrose and α,α-Trehalose in Aqueous Solutions. *J. Phys. Chem. A* **2003**, *107*, 936–943.

(11) Sartorio, R.; Wurzburger, S.; Guarino, G.; Borriello, G. Diffusion of polyols in aqueous solution. *J. Solution. Chem.* **1986**, *15*, 1041–1049.

(12) Hatcher, E. R.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates, and Inositol. *J. Chem. Theory Comput.* **2009**, *5*, 1315–1327.

(13) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(14) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(15) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(16) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(17) 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.

(18) Hess, B. Determining the shear viscosity of model liquids from molecular dynamics simulations. *J. Chem. Phys.* **2002**, *116*, 209–217.

(19) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci.* **2001**, *98*, 5578–5583.

(20) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71 – 73.

(21) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(22) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, 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.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.