

Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States

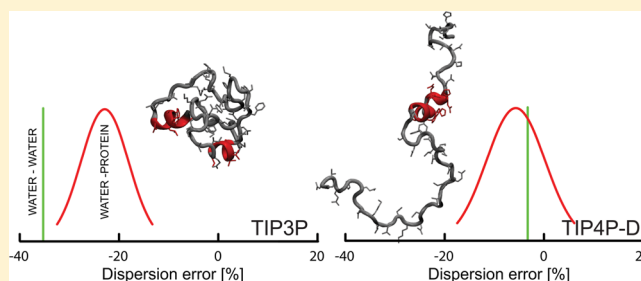
Stefano Piana,^{*,†} Alexander G. Donchev,[†] Paul Robustelli,[†] and David E. Shaw^{*,†,‡}

[†]D. E. Shaw Research, New York, New York 10036, United States

[‡]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

S Supporting Information

ABSTRACT: Many proteins can be partially or completely disordered under physiological conditions. Structural characterization of these disordered states using experimental methods can be challenging, since they are composed of a structurally heterogeneous ensemble of conformations rather than a single dominant conformation. Molecular dynamics (MD) simulations should in principle provide an ideal tool for elucidating the composition and behavior of disordered states at an atomic level of detail. Unfortunately, MD simulations using current physics-based models tend to produce disordered-state ensembles that are structurally too compact relative to experiments. We find that the water models typically used in MD simulations significantly underestimate London dispersion interactions, and speculate that this may be a possible reason for these erroneous results. To test this hypothesis, we create a new water model, TIP4P-D, that approximately corrects for these deficiencies in modeling water dispersion interactions while maintaining compatibility with existing physics-based models. We show that simulations of solvated proteins using this new water model typically result in disordered states that are substantially more expanded and in better agreement with experiment. These results represent a significant step toward extending the range of applicability of MD simulations to include the study of (partially or fully) disordered protein states.



INTRODUCTION

The field of structural biology has for the most part been dominated by the study of the folded states of proteins. Most folded states are characterized by a well-defined three-dimensional structure, which can often be determined to atomistic resolution using conventional experimental techniques like X-ray diffraction or nuclear magnetic resonance (NMR) spectroscopy. Recently, however, it has been recognized that many proteins that perform important biological functions can populate partially or completely disordered states under physiological conditions. These disordered states are typically composed of a heterogeneous ensemble of rapidly interconverting conformations.

It is very challenging to experimentally characterize a conformational ensemble of disordered states with the same level of resolution that can be achieved for the folded states of proteins, since disordered states are not amenable to crystallization, and NMR-derived data may not be sufficient to determine the structural properties of such a heterogeneous ensemble with atomistic resolution. Molecular dynamics (MD) simulations, on the other hand, generate continuous, atomistically detailed trajectories that might be expected to provide a valuable complement to experimental data in characterizing the structural and dynamic properties of disordered states.

Unfortunately, there is evidence suggesting that, at least in some instances, the underlying physical models (“force fields”) typically used in MD simulations may grossly misrepresent the

conformational ensemble of the disordered states of proteins.^{1–4} In previous MD simulation studies,^{5–10} for example, it has been observed that the expanded states of a disordered ensemble are often unstable in simulation and tend to collapse to so-called “molten globule” states. The structures that comprise these molten globule states are as compact as the structures that represent the folded states of comparably sized proteins,^{11–13} a result which disagrees with experimental observations.¹⁴ Such compact molten globule states sometimes display other unusual properties, such as very slow relaxation times and abnormally high degrees of secondary structure content.^{7,15}

Although many of the results obtained from MD simulations of disordered proteins clearly deviate from experiment, the cause and extent of the force field deficiencies that underlie these disparities remain unclear. If widespread, such deficiencies could seriously hamper the systematic use of MD simulation in characterizing the disordered states of proteins. To ameliorate these problems, simulations have sometimes been performed under nonphysical or nonphysiological conditions that tend to counter some of the more evident discrepancies observed in disordered-state simulations; examples include the use of implicit solvent at high temperature^{13,16,17} and the introduction

Received: September 4, 2014

Revised: March 11, 2015

Published: March 12, 2015

of restraints based on experimental data to either drive the conformational sampling¹⁸ or select a subset of conformations that are consistent with the experiment.¹⁹ These methods, however, may not be sufficiently accurate for all systems.

In this paper, we begin by systematically assessing the ability of a number of commonly used protein force fields and water models to reproduce various structural properties of disordered proteins under physiological conditions. We find that all combinations of force fields and water models tested result in disordered states that are overly compact, with radii of gyration (R_g) that are in strong disagreement with experimental observations. We conjecture that an important reason for the discrepancies between these simulated and experimental results is that current water models severely underestimate water–water and water–solute dispersion interactions.^{2,20,21} We evaluate this hypothesis by performing high-level quantum mechanics (QM) calculations. In agreement with earlier results,²² our calculations indeed indicate that all tested water models fail to capture the full extent of dispersion stabilization energy. Remarkably, this discrepancy is largest for those intermolecular geometries corresponding to the most stable, hydrogen-bonded conformations.

We then introduce a new, four-point water model, called TIP4P-D, in which the water dispersion coefficient C_6 is constrained to be ~50% larger than in current water models, and the remaining nonbonded parameters are optimized with respect to experimental liquid water properties. We find that, even while imposing this constraint on the value of the water dispersion coefficient, it is possible to obtain a water model with good liquid water properties. Simulations performed with this new water model, featuring increased dispersion, result in disordered states that are substantially more expanded and exhibit good quantitative agreement with most of the available experimental data.

METHODS

MD Simulations of Disordered States of Proteins. We performed simulations of five proteins: the apo N-terminal zinc-binding domain of HIV-1 integrase (IN) (PDB entry 1WJB),²³ the immunoglobulin-binding domain of protein L (PDB entry 2PTL),²⁴ the cold-shock protein from *Thermotoga maritima* (CspTm) (PDB entry 1G6P),²⁵ α -synuclein (PDB entry 1XQ8),²⁶ and prothymosin- α . These proteins were chosen, as there is a large amount of experimental data characterizing their unfolded states under physiological conditions. Simulations were started from an extended conformation. Initial extended structures of IN, protein L, and CspTm were solvated in cubic $70 \times 70 \times 70 \text{ \AA}^3$ boxes, containing ~10 000 water molecules and 0.1 M NaCl, whereas the larger α -synuclein and prothymosin- α were solvated in an $\sim 100 \times 100 \times 100 \text{ \AA}^3$ box containing ~40 000 water molecules and 0.1 M NaCl. The CHARMM22 force field²⁷ was used to represent the ions. The systems were initially equilibrated at 300 K and 1 bar for 1 ns using the Desmond software;²⁸ production runs at 300 K were performed in the NPT ensemble^{29–31} with the Anton specialized hardware³² using a 2.5 fs time step and a 1:2 RESPA scheme.³³ Bonds involving hydrogen atoms were restrained to their equilibrium lengths using the M-SHAKE algorithm.³⁴ Nonbonded interactions were truncated at 12 \AA , and the Gaussian split Ewald method³⁵ with a $32 \times 32 \times 32$ mesh ($64 \times 64 \times 64$ for α -synuclein and prothymosin- α) was used to account for the long-range part of the electrostatic interactions.

MD simulations of IN were performed using the Amber12,³⁶ the Amber99SB-ILDN,^{37,38} the Amber03*,³⁹ the CHARMM22*,⁴⁰ and the OPLS^{41,42} force fields using either the TIP3P,⁴³ the TIP4P-EW,⁴⁴ the TIP4P/2005,⁴⁵ or the TIP4P-D water models. MD simulations of protein L, CspTm, α -synuclein, and prothymosin- α were performed using the CHARMM22*, the Amber99SB-ILDN, and the Amber12 force fields⁴⁶ and either the TIP3P, the TIP4P-EW, or the TIP4P-D water models. In most simulations, relaxation from the initial conditions required several hundred ns to a few μs , and each simulation was run for a minimum of 5 μs . To fully verify the convergence of R_g , several simulations were extended up to 130 μs . We found that it typically takes a few μs of simulation to establish whether or not a force field has a tendency to form overly collapsed or more expanded disordered states and to obtain a rough estimate of R_g . On the other hand, substantially longer time scales (from 10 to 100 μs) can be required to quantitatively estimate other structural properties, like the secondary structure content or the distance distribution between two residues (Figure S1, Supporting Information), in agreement with previous findings.⁴⁷

A total of 41 simulations, with an aggregate length of ~830 μs , were performed; a detailed list is reported in Table S1 (Supporting Information). The average R_g was calculated from the squared average of the R_g values of the protein atoms obtained in simulation. Standard errors of the mean were estimated using a blocking procedure.⁴⁸

Ab Initio Calculation of Dispersion Interactions.

Conventionally, the dispersion component of the intermolecular interaction energy (E_{DS}) is represented as an inverse power series of the molecule–molecule separation r

$$E_{\text{DS}} = - \sum_{n=6,8,10,\dots} \frac{C_n}{r^n} \quad (1)$$

where the C_n are positive-valued dispersion coefficients that can be obtained from frequency-dependent molecular polarizabilities using the Casimir–Polder integral.⁴⁹ Strictly speaking, this approach is accurate only in the asymptotic regime, in which r significantly exceeds the molecular sizes. At shorter separations, one should also consider the anisotropic contributions to E_{DS} . Finally, when there is non-negligible overlap between molecular electronic densities, this approximation fails unless a damping correction, for which a canonical form has not been firmly established, is applied to every term in eq 1.

A more practical approach to obtain E_{DS} is to decompose the total intermolecular interaction energy computed at a sufficiently high level of quantum theory (i.e., one that includes a reliable representation of dispersion effects) into its principal components: electrostatics, exchange repulsion, induction, and dispersion. Following the decomposition scheme presented in ref 50, we define E_{DS} as the post-Hartree–Fock correlation interaction energy minus the correlation corrections to the electrostatic and exchange-repulsion components. The correlation energy is estimated at the CCSD(T) level of theory with the aug-cc-pVTZ(-f) basis set. It is further corrected for the basis-set incompleteness effects, estimated as the difference between the aug-cc-pVQZ and aug-cc-pVTZ(-f) correlation energy at the MP2 level of theory. Such an approach was found to be in very good agreement with earlier results for dispersion energies computed at dimer equilibrium geometries,²² as well as with the computationally more demanding symmetry-adapted perturbation theory (SAPT) approach⁵¹ (with a typical

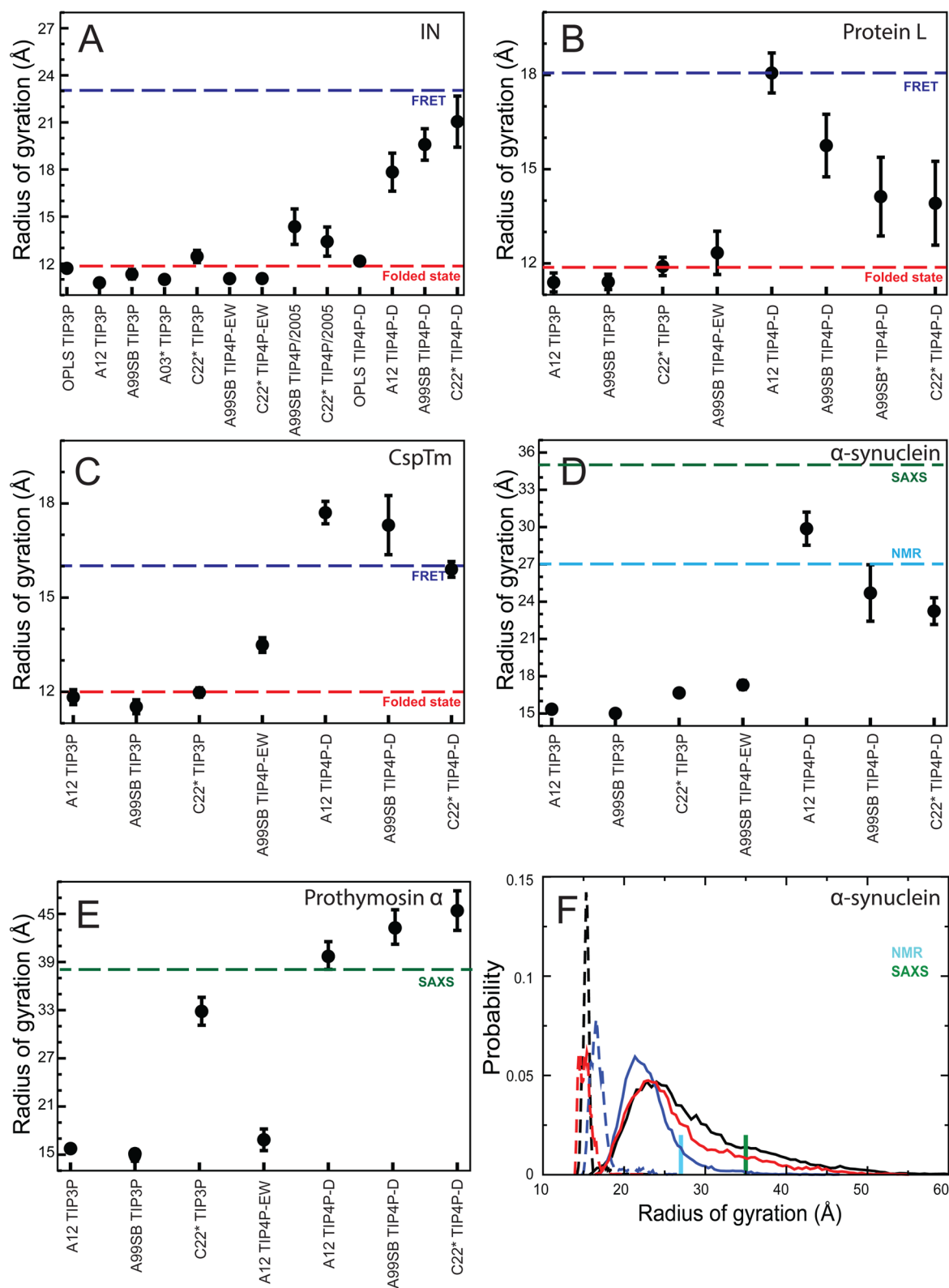


Figure 1. R_g of disordered proteins from simulation. (A) R_g of the natively unfolded IN calculated from simulations performed with the OPLS,^{41,42} CHARMM22* (C22*),⁴⁰ Amber03* (A03*),³⁹ Amber99SB-ILDN (A99SB),³⁷ and Amber12 (A12)³⁶ force fields using either the TIP3P,⁴³ the TIP4P-EW,⁴⁴ the TIP4P/2005,⁴⁵ or the TIP4P-D water model. By way of comparison, we report the R_g of the zinc-bound folded protein (PDB entry 1WJB; red) and the R_g of the unfolded state as determined from FRET (blue)^{17,60} experiments. The same plot is also reported for protein L with FRET estimates of the R_g at zero denaturant⁹² (B), CspTm¹⁷ (C), α -synuclein^{60,73,74} (D), and prothymosin- α ⁹³ (E). (F) The distribution of the radius of gyration observed for α -synuclein in simulations performed with Amber99SB-ILDN (blue), CHARMM22* (black), and Amber12 (red) and either the TIP3P (dashed line) or the TIP4P-D (solid line) water models. Estimates of R_g obtained experimentally using SAXS⁷⁴ and NMR⁷³ are also indicated.

Table 1. Parameters and Physical Properties of Selected Commonly Used Water Models and TIP4P-D^a

	Expt	TIP3P	SPC/E	TIP4P-EW	TIP4P/2005	TIP4P-D
μ (D)	>2.6 ^{69,70}	2.35	2.35	2.32	2.305	2.403
hydrogen charge		0.417	0.4238	0.52422	0.5564	0.58
C_6 (kcal mol ⁻¹ Å ⁶)	622 ^{63,64}	595	625	653	736	900
C_{12} (kcal mol ⁻¹ Å ¹²)		582.000	629.482	656.138	731.380	904.657
ΔH_v (kcal mol ⁻¹)	10.5	10.2	10.5	10.6	11.0	11.3
C_p (cal mol ⁻¹ K ⁻¹)	18.0	15.2	17.3	17.6	17.8	16.8
T_{md} (K)	277		250	270	275	270
ρ (T_{md}) (g cm ⁻³)	1.000		1.012	1.000	1.001	0.997
γ (mN m ⁻¹)	71.7	47.8	58.4	59.2	63.3	71.2
α_v (10 ⁻⁴ K ⁻¹)	2.53	8.5	4.7	3.1	2.7	2.6
D (10 ⁻⁵ cm ² s ⁻¹)	2.3	5.8	2.6	2.6	2.2	2.1
ϵ_0	78	96(3)	72(4)	63(3)	56(2)	68(2)

^aThe parameters of the water models (dipole moment, μ , and LJ parameters C_6 and C_{12}) and theoretical estimates for the dispersion coefficient C_6 and the dipole moment of a molecule of water in liquid water (μ). Also shown are the vaporization enthalpy (ΔH_v), heat capacity (C_p), temperature of maximum density (T_{md}), density at T_{md} (ρ), surface tension (γ), isothermal compressibility (α_v), diffusion coefficients (D), and dielectric constant (ϵ_0) at 300 K and 1 bar, unless otherwise indicated. Vaporization enthalpy values ΔH_v for all models except TIP3P include a self-polarization correction.⁵⁵ The heat capacity values include vibrational corrections.⁴⁴

difference of only a few percent). All quantum calculations presented here were performed within the MOLPRO^{52,53} package.

Development of the TIP4P-D Water Model. The development of a new four-point water model (TIP4P-D) is presented as a proof of principle to demonstrate the importance of water dispersion interactions in the simulation of the disordered states of proteins. In the parametrization of TIP4P-D, the geometry of the TIP4P/2005 water model⁴⁵ was retained. To capture a larger fraction of E_{DS} , we first constrained the dispersion coefficient C_6 to a value 50% larger than in the TIP3P water model (for further discussion, see “TIP4P-D, a water model with increased dispersion interactions” in the Results and Discussion section) while retaining the simple form of the widely used 12-6 Lennard-Jones (LJ) potential. Charges and the C_{12} parameters were then adjusted to fit density and vaporization enthalpy temperature profiles.

Liquid water properties were obtained from MD simulations of a $45 \times 45 \times 45$ Å³ box containing 3054 water molecules. Simulations were run for 5 ns at temperatures ranging from 240 to 390 K using the particle mesh Ewald (PME) method (to account for long-range electrostatic interactions),⁵⁴ isotropic long-range corrections for the LJ interactions, and a 10 Å cutoff. A self-polarization correction⁵⁵ was included in the vaporization enthalpy calculation. Surface tension was calculated with the virial method⁵⁶ using a $45 \times 45 \times 90$ Å³ box containing 3054 water molecules. Simulations of 10 ns were performed in the NVT ensemble for each water model using a 12 Å cutoff and no long-range tail corrections to the LJ interactions.

Hydration free energies of small molecules representing the side chains of the amino acids that are not charged at neutral pH were calculated using the free-energy perturbation approach and the Desmond software.²⁸ Each molecule was immersed in a $30 \times 30 \times 30$ Å³ box containing ~800 water molecules. A cutoff of 10 Å was used for the electrostatic and LJ interactions. The PME method⁵⁴ was used to account for the long-range electrostatic interactions. A time step of 2.0 fs was used to integrate the equations of motion. Interactions between the solute and the solvent were switched off in 16 λ -windows, with 1.5 ns of simulation performed for each λ -window. Hydration free energies were calculated using the Bennett acceptance ratio.⁵⁷

RESULTS AND DISCUSSION

Assessment of the Ability of Current Force Fields to Describe the Disordered States of Proteins. To systematically assess the ability of simulations to reproduce disordered states, we performed simulations of the apo form of the N-terminal zinc-binding domain of HIV-1 integrase (IN)²³ using a number of common force fields and solvent models. IN is a zinc-binding protein that is natively unfolded under standard conditions, as long as zinc is removed from the solution.⁵⁸ The R_g of the unfolded state of IN has been determined using Förster resonance energy transfer (FRET),⁵⁹ and the estimated value of 24 Å⁶⁰ indicates that it should be substantially more expanded than the folded state (which has an R_g value of ~12 Å when the flexible terminal residues are omitted).

Simulations of IN performed with representatives of the most widely used biomolecular force fields (OPLS, CHARMM22*, Amber03, and Amber99SB-ILDN) and the TIP3P water model result in an unfolded state with an R_g value of ~10–12 Å (i.e., roughly as compact as the folded state (Figure 1A) and in strong disagreement with the FRET experimental data). The use of the Amber99SB force field with a four-point water model has been recommended for the simulation of disordered proteins.^{1,61} We only observed a modest increase in the R_g value of the unfolded state in Amber and CHARMM simulations performed with the TIP4P-EW or TIP4P/2005 water models (Figure 1A).

These findings do not appear to be specific to IN, as similar results were also obtained for the other four proteins investigated using a smaller subset of force fields and water models: CspTm (Figure 1B), protein L (Figure 1C), α -synuclein (Figure 1D), and prothymosin- α (Figure 1E). Only a simulation of the highly charged prothymosin- α performed using CHARMM22* resulted in somewhat expanded disordered states. In all other cases, the disordered states observed in simulations are much more collapsed than those observed experimentally, and for CspTm and protein L, they are as compact as the folded state (whose R_g can be taken as a reference value for a maximally compact chain). This result strongly suggests that these overly compact disordered states are a general feature of current force fields.

Comparison of Force-Field-Based Dispersion Interactions to High-Level Quantum Data. The observation that

different combinations of force fields and water models all give similar results in terms of the collapse of the protein chain for disordered proteins prompted us to seek common features in these simulations that might explain this finding. We noted that all water models commonly used in biomolecular simulations span a relatively small area of parameter space and feature very similar values for C_6 (Table 1). A possible explanation for this finding is that, in particular for early water parametrizations, the fit to the experimental data typically targeted by the developers was not strongly influenced by the water dispersion coefficient.^{20,62} Parametrizations that resulted in C_6 values close to the theoretical estimate based on QM calculations ($\sim 600 \text{ kcal mol}^{-1} \text{ \AA}^6$)^{63–65} were generally considered acceptable.^{43,66} No water model currently used for protein simulations deviates much from this value (Table 1). The theoretical estimate, however, only describes the long-range component of E_{DS} . At the typical nearest-neighbor distances ($\sim 3 \text{ \AA}$ between the oxygen atoms) observed in liquid water, the cumulative effect of the higher-order C_8/r^8 and C_{10}/r^{10} terms is no longer a small correction and can even exceed the C_6/r^6 contribution to E_{DS} .⁶⁴ This observation led us to suspect that current water models may severely underestimate the full extent of dispersion interactions in the liquid phase.

To verify whether this is indeed the case, we compared the dispersion component of the intermolecular interaction energy obtained at a high level of quantum theory (see the Methods section) to calculations using three popular force fields: OPLS, Amber, and CHARMM (in conjunction with their typically prescribed water models). Figure 2 shows the results of this comparison for a range of dimers of small molecules representative of the functional groups found in proteins (see more details in Table S3, Supporting Information). Despite having long tails, all the three error distributions corresponding to the interactions not involving water feature a relatively small systematic bias, indicating that, on average, these interactions are reasonably balanced. On the other hand, interactions that involve water are systematically underestimated by the three force fields, with the largest discrepancies observed for the water–water dimers. Figure S2 (Supporting Information) provides a more detailed picture by comparing, across a range of intermolecular separations, the water dimer results with the results for two nonpolar examples: butane and benzene dimers; the error is largest at the water–water equilibrium distance, where the interaction is underestimated by roughly a factor of 2.

This observation that water dispersion interactions are severely underestimated is applicable beyond the three water models considered here, as nearly all popular water models feature very similar C_6 coefficients (Table 1). This finding calls into question whether current water models correctly describe the relative balance between the highly directional hydrogen-bonding interactions and the nearly isotropic dispersion interactions. Obtaining the correct balance between these interactions is expected to be an important factor in accurately describing the solvation of proteins and may have important consequences for the ability of simulations to correctly describe the highly hydrated disordered states.

TIP4P-D, a Water Model with Increased Dispersion Interactions. To demonstrate the importance of achieving a better balance between water dispersion and electrostatic interactions for biomolecular simulations, we parametrized a new water model (TIP4P-D) that features a significantly higher dispersion coefficient: $C_6 = 900 \text{ kcal mol}^{-1} \text{ \AA}^6$ (i.e., $\sim 50\%$ larger

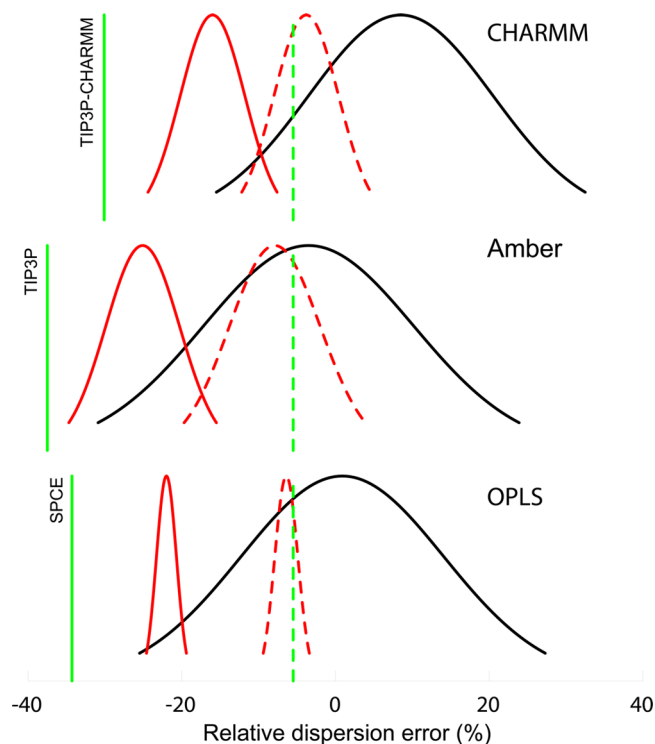


Figure 2. Dispersion energy comparison between high-level quantum calculations and three popular force fields: OPLS/SPCE (lower panel), Amber/TIP3P (middle panel), and CHARMM/TIP3P-CHARMM (upper panel). The comparison was performed for a diverse set of dimers of small molecules representative of the functional groups found in proteins. Shown are Gaussian approximations to the actual error distributions. The data set is split into three categories: water dimer (green vertical lines), water interacting with other molecules (red curves), and interactions not involving water (black curves). Dashed lines correspond to the TIP4P-D water model. The vertical axis is in arbitrary units. All distributions are normalized to the same value at the maximum.

than in most popular models; see Table 1). It is important to stress that the C_6/r^6 functional form is intrinsically incapable of accurately reproducing E_{DS} at all dimer separations (Figure S2, Supporting Information). On the other hand, using a more realistic functional form (e.g., explicitly considering the additional C_8/r^8 and C_{10}/r^{10} terms) would not allow us to preserve compatibility with standard biomolecular force fields, which describe dispersion interactions using only the C_6/r^6 term. Our choice of the C_6 value is thus an attempt to strike a balance between compensating for the missing higher-order (r^{-8} , r^{-10} , etc.) terms and keeping the asymptotic error reasonably small. This results in short-range dispersion interaction energies similar to those of the highly accurate, polarizable iAMOEBA water model,⁶⁷ which features a more flexible functional form⁶⁸ for the dispersion interaction.

Imposing an additional constraint on the C_6 coefficient naturally limits the model's ability to reproduce experimental liquid water properties; nevertheless, as can be seen in Table 1, this unusually high C_6 value did not prevent TIP4P-D from reproducing most of the key liquid water properties (including the diffusion and surface tension, which were not in our training set) at a level of accuracy similar to other four-point water models. It is interesting to note that, along with enhanced dispersion, TIP4P-D features slightly stronger electrostatic interactions, as represented by the dipole moment, which is

0.05–0.1 D larger than that of any other model in Table 1, and is in slightly better agreement with *ab initio* simulations.^{69,70} A larger dipole moment is necessary to reproduce with reasonable accuracy the temperature of maximum density and the thermal expansion coefficient but leads to an overestimation of the vaporization enthalpy.

It has been shown that the hydration free energies of small molecules can be strongly influenced by changes in the water-solute dispersion coefficient.²¹ Using the TIP4P-D water model, we calculated the hydration free energies ΔG_h for small molecules representing the side-chain groups of natural amino acids and compared them with the ΔG_h values obtained using standard water models (Figure 3). The comparison shows

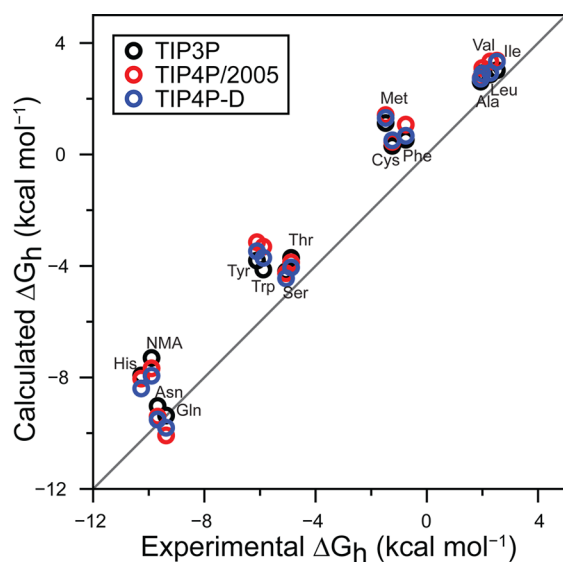


Figure 3. Solubility of small molecules in different solvent models. Hydration free energy (ΔG_h) of small molecular fragments in the TIP3P (black circles), TIP4P/2005 (red circles), and TIP4P-D (blue circles) water models compared to the values determined experimentally. Each fragment is a model for an amino acid side chain; the name of the corresponding amino acid is reported near each data point. *N*-Methylacetamide (NMA) is the small molecule used as a model for the protein backbone. The Amber99SB force field was used to describe the molecular fragments. A very similar plot is obtained for the CHARMM22 force field (see Figure S4, Supporting Information).

that the modified water model improves agreement with experimental values of ΔG_h , though only marginally. This suggests that, in this case, the difference in dispersion coefficient does not strongly influence the hydration properties of small molecules.

Determining the Effect of Increasing Water Dispersion on Simulations of Disordered Protein Ensembles. While it has been proposed² that many liquid water properties are relatively insensitive to the exact value of the dispersion coefficient, a proposition that largely agrees with our findings, some effect is observed on surface tension, which appears to grow as the water dispersion coefficient (C_6) increases⁷¹ (Table 1). This observation suggests that the hydration and structural properties of large solutes may be affected, as they are also influenced by water surface tension.⁷² To examine how a change in dispersion influences the results of protein simulations, we compare protein simulations performed using the TIP4P-D model with simulations performed using

conventional water models, with a particular emphasis on the properties of disordered states.

We find that, for all five disordered proteins considered (IN, CspTm, protein L, α -synuclein, and prothymosin- α), Amber and CHARMM simulations performed with TIP4P-D result in disordered states that are substantially more expanded and that are generally in much better agreement with the experimental estimates than those obtained with any combination of the original force fields (Figure 1A–E). It is notable that this effect is observed despite the fact that the solubility of the individual protein fragments in isolation, as inferred by hydration free energy calculations performed on small-molecule analogues, is not strongly influenced by the choice of water model (Figure 3).

Although the effect of TIP4P-D water on the size of disordered states is fairly general, not all force fields behave in exactly the same way. OPLS simulations with TIP4P-D result in only marginal improvements in R_g (Figure 1A), suggesting that this force field may require even larger values of the water dispersion coefficient to transition to more expanded states. On the other hand, simulations of the highly charged prothymosin- α performed with CHARMM22* and the TIP3P water model result in reasonably expanded disordered states, suggesting that this force field may be suitable for simulations of highly charged disordered systems even in conjunction with standard water models. In this respect, it is interesting to note that, among the three force fields shown in Figure 2, CHARMM has the smallest systematic error for the interactions involving water.

These results strongly suggest that the dispersion interactions are an important component for determining the hydration properties of large macromolecules, and in particular for fine-tuning the balance between disordered, expanded states and compact, collapsed states.

Detailed Comparison of the α -Synuclein Simulations to Experimental Data. A large amount of experimental data is available for α -synuclein, allowing a robust assessment of the quality of the structural ensembles observed in simulations. Here we compare the Amber12 results obtained in TIP3P and TIP4P-D simulations with NMR, small-angle X-ray scattering (SAXS), and FRET data that probe both the local and nonlocal structural and dynamic properties of the ensemble (Figures 4 and 5). The simulation run using the TIP4P-D water model is in substantially better agreement with experimental measurements for nearly all of the data presented here than the simulation run with TIP3P.

The average R_g observed in the simulation using TIP4P-D (29 Å) is much larger than that observed in TIP3P (15 Å) and is in much closer agreement with the 27–35 Å experimental estimates obtained from NMR⁷³ and SAXS⁷⁴ measurements. Accordingly, the average SAXS profile, calculated with the program FoXS,^{75,76} from the TIP4P-D simulation is in much better agreement with the experimental profile⁷⁴ than the profile obtained from the TIP3P simulation (Figure 4A).

NMR paramagnetic relaxation enhancement (PRE) measurements provide detailed information about the relative populations of long-range contacts. We have compared the simulation results with a large set of previously measured intramolecular PRE data obtained from 12 paramagnetic labels containing information on 1226 interatomic distances.^{77–81} Two representative PRE intensity-ratio profiles are displayed in Figure 4B, and the full comparison to all 12 spin-label profiles is reported in Figure S3 (Supporting Information). The TIP3P simulation contains too many persistent long-range contacts

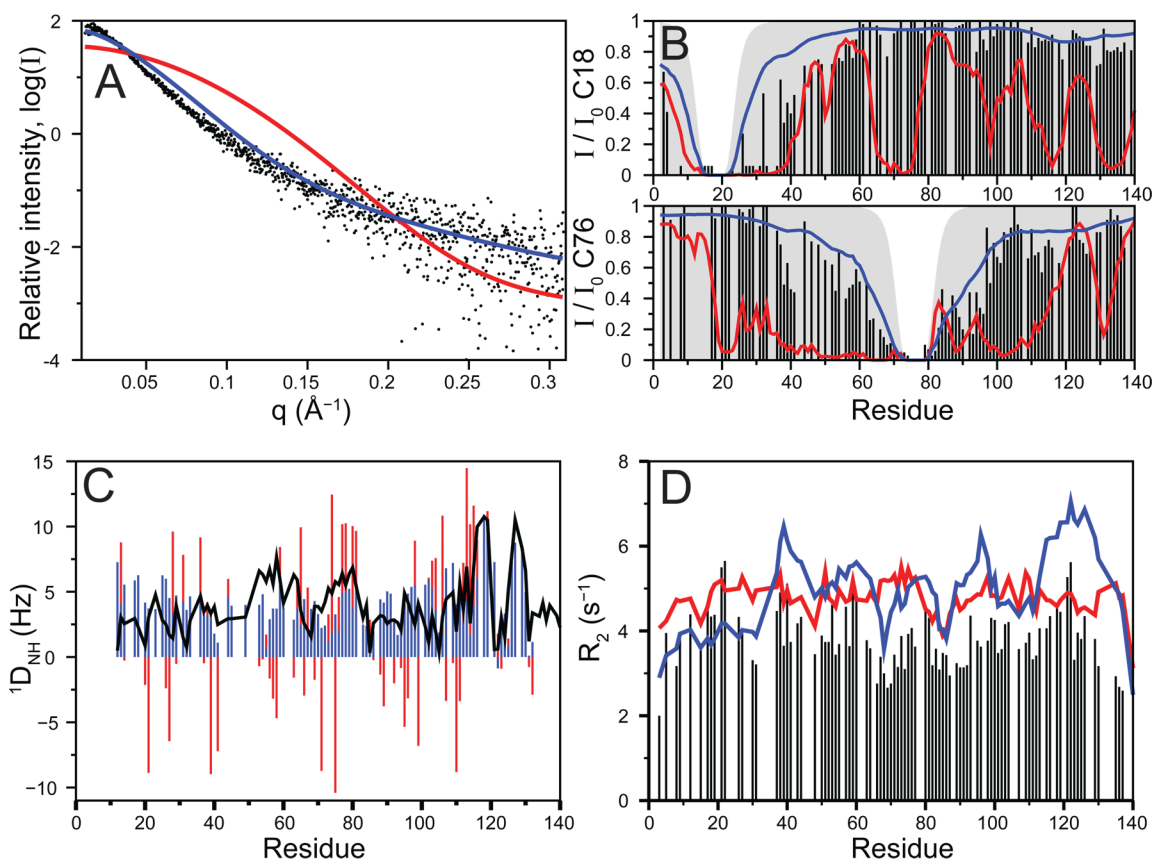


Figure 4. Reproduction of experimental α -synuclein data from simulations. Comparison of calculated and experimental SAXS and NMR observables for α -synuclein simulations performed in TIP3P (red) or TIP4P-D (blue) water with the Amber12 force field. Experimental values are displayed in black. (A) SAXS curve⁷⁴ calculated with the program FoXS.^{75,76} (B) PRE intensity ratios from representative spin labels placed at residues 18 and 76.^{78,79} The shaded gray area represents the expected PRE values for a random coil. (C) $^1D_{NH}$ RDCs⁷⁸ calculated with the program PALES⁹⁴ using a local alignment window of 15 residues. (D) ^{15}N R_2 relaxation rates⁸⁰ calculated from simulated spectral densities.⁴⁷

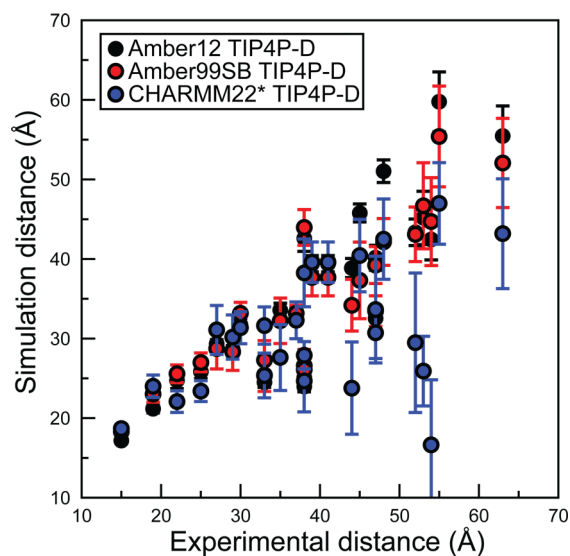


Figure 5. Inter-residue distances in α -synuclein. Comparison of simulated and experimental^{61,86,87} mean distances between residues in α -synuclein. Distances in simulation were calculated as the average $C_\alpha-C_\alpha$ distance over 20 μs of simulation performed in TIP4P-D water with the Amber12 force field (black), the Amber99SB-ILDN force field (red), or the CHARMM22* force field (blue). The first 0.5 μs of simulation were not considered in the analysis.

compared to the experimental PREs because the sampled structures are too compact. Both in the experiments and in the TIP4P-D simulation, the population of intramolecular contacts deviates from the behavior expected from a random coil, though contacts that are distant in sequence are often populated to a lesser extent in the simulation than in experiment. These results suggest that the sampling of extended structures in TIP4P-D simulations is not achieved at the cost of a complete elimination of experimentally observed long-range contacts.

NMR chemical shifts and residual dipolar couplings (RDCs) are sensitive probes of the local conformational preferences of intrinsically disordered proteins.^{82–84} We observe significant improvements for chemical shifts and RDCs back-calculated from the structures sampled in the TIP4P-D simulation compared to the TIP3P ensemble. The deviation between the calculated $^1D_{NH}$ RDCs and the experimental values, quantified by the Q -factor, is 0.51 for the TIP4P-D simulation and 1.26 for TIP3P (Figure 4C). The average root-mean-square deviations (RMSDs) between calculated and experimental chemical shifts in the TIP4P-D simulation are 0.63, 0.19, 1.95, 0.34, and 0.92 ppm (ppm) for C_α , HN, N, C' , and C_β shifts, respectively; these deviations are substantially smaller than the values of 1.18, 0.38, 3.26, 0.67, and 1.22 ppm that were obtained for the TIP3P simulation.

Importantly, the chemical shift prediction errors obtained for the TIP4P-D ensemble are lower than the average prediction errors obtained from a benchmark database of static X-ray

structures of globular proteins,⁸⁴ and the prediction errors of the TIP4P-D ensemble are similar in accuracy to the errors obtained by sequence-based predictions of random-coil chemical shifts.⁸⁵ Taken together, the agreement between calculated and experimental chemical shifts and RDCs suggests that the TIP4P-D simulation provides a reasonably accurate description of the local conformational properties of α -synuclein.

NMR ¹⁵N R_2 transverse relaxation rates reflect the frequency distributions of the motions of amide bond vectors in proteins. R_2 values calculated from MD simulations have previously been demonstrated to be strongly affected by the structural features of intrinsically disordered protein ensembles, the time scales of intramolecular motions, and the time scales of overall rotational tumbling.^{47,83} We compared R_2 relaxation rates calculated from simulation runs using TIP4P-D and TIP3P to the experimental values⁸⁰ (Figure 4D). The R_2 rates back-calculated from the simulation run with TIP3P have a somewhat lower RMSD from the experimental values (1.16 Hz) than does the simulation run using TIP4P-D (1.42 Hz). They show, however, smaller sequence variations and have a lower correlation coefficient with the experimental values ($r = 0.24$) compared to the simulation run using TIP4P-D ($r = 0.40$).

These results suggest that, for the molten globule-like structures observed in the TIP3P simulation, relaxation is largely determined by the overall rotational tumbling, which affects all residues similarly, and that R_2 rates do not contain useful information on local flexibility. On the contrary, the fluctuations of the R_2 rates calculated from TIP4P-D simulations, although globally shifted toward values that are slightly too high, correctly identify the regions of higher R_2 rates centered near residues 40, 60, 80, 100, and 120, which suggests that the fluctuations provide a meaningful measure of local flexibility. We note that, if the time axis of the amide bond vector autocorrelation functions used to calculate the R_2 rates is scaled to correct for the slightly excessive viscosity of the TIP4P-D water model relative to experiment, the agreement with experimental R_2 rates improves (RMSD = 1.14 Hz) but remains globally shifted to values that are too high. (The average simulated R_2 is 4.76 Hz, compared to the experimental average of 3.86 Hz.) This suggests that the description of the rotational and segmental diffusion of proteins in TIP4P-D water could be further improved.

Finally, a number of intraresidue distances have been measured experimentally for this protein using fluorescence spectroscopy techniques.^{61,86,87} Although long-range distances calculated from TIP3P simulations are invariably much shorter than the experimental estimates, the Amber12 and Amber99SB simulations performed with TIP4P-D water reproduce the experimental results quite well (Figure 5). Simulations performed with the CHARMM22* force field in TIP4P-D water are in somewhat worse agreement with the experimental data (Figure 5), with some long-range contacts having a high probability to form. Analysis of the trajectory suggests that the underlying reason is the repeated formation of knotted structures that persist on the 1–10 μ s time scale during the simulation. This results in a structural ensemble with persistent long-range contacts that is very different from the ensembles of the Amber simulations, where no knotting occurs.

Simulation of Folded Proteins Using the TIP4P-D Water Model. The dramatic effect of the TIP4P-D water model on simulations of the disordered states of proteins raises the legitimate concern that it could excessively destabilize

simulations of the ordered, folded states.²¹ To address this question, we performed 10 μ s simulations of the native states of ubiquitin and GB3 in TIP4P-D water. Both proteins were stable on the 10 μ s time scale, but the simulations resulted in structural ensembles that were more dynamic than those obtained with standard water models and were in somewhat worse agreement with NMR data (Table S2, Supporting Information), suggesting that the TIP4P-D water model may slightly destabilize the folded states of proteins. We quantified the amount of destabilization by performing reversible-folding simulations of a WW domain (using Amber99SB-ILDN) and of the villin headpiece (using Amber12) in TIP4P-D water. These simulations indicate that, at around the melting temperature, the native states of these two proteins are destabilized by ~ 2 kcal mol⁻¹ in TIP4P-D compared to TIP3P.

We speculate that, using standard water models, the strong tendency to produce collapsed structures may lead to an overstabilization of the compact folded state that could mask other force field deficiencies. Because the TIP4P-D water model strongly reduces the propensity for collapse, the effects of other force field inaccuracies are more evident. Exposing these effectively “hidden” force field problems should allow researchers to address them more directly, which may ultimately lead to more accurate models and improved simulations of folded proteins.

CONCLUSION

Simulations based on current physics-based water models and protein force fields systematically fail to describe key structural properties of disordered states—a finding we attribute to a failure of commonly used water models to account for the full extent of dispersion interactions. We have created a new water model, TIP4P-D, that retains the standard 12-6 LJ potential (thus preserving compatibility with most existing biomolecular force fields) while increasing the strength of dispersion interaction. Simulations performed with TIP4P-D result in a slight destabilization of native states, but the model greatly improves the description of disordered protein states, allowing meaningful comparisons to be made between the results of unrestrained atomistic simulations and experimental studies of such states. These results are very encouraging, as we (i) did not exhaustively search for an optimal value for the water dispersion coefficient and (ii) did not modify parameters of the protein force fields. We expect that the agreement between experiments and simulations for various protein states could be further improved by using more realistic functional forms. This would, however, require redesigning and reparameterizing not only the water model but also the protein force field.

ASSOCIATED CONTENT

Supporting Information

Further discussion of the parametrization of water models for biomolecular simulations; additional Methods information; Tables S1–S3 and Figures S1–S4. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: Stefano.Piana-Agostinetti@DEShawResearch.com. Phone: (212) 403-8165. Fax: (646) 873-2165.

*E-mail: David.Shaw@DEShawResearch.com. Phone: (212) 478-0260. Fax: (212) 845-1286.

Notes

The authors declare the following competing financial interest(s): This study was conducted and funded internally by D. E. Shaw Research, of which D.E.S. is the sole beneficial owner and Chief Scientist, and with which all authors are affiliated.

ACKNOWLEDGMENTS

The authors thank Andrew Taube for help with and discussion of QM issues, John Klepeis for a critical reading of the manuscript, Berkman Frank for editorial assistance, and Carlos Bertocini, David Eliezer, Xavier Salvatella, and Markus Zweckstetter for sharing data on α -synuclein.

REFERENCES

- (1) Johnson, M. E.; Malardier-Jugroot, C.; Murarka, R. K.; Head-Gordon, T. Hydration Water Dynamics near Biological Interfaces. *J. Phys. Chem. B* **2009**, *113* (13), 4082–4092.
- (2) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122* (13), 134508.
- (3) Demerdash, O.; Yap, E. H.; Head-Gordon, T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu. Rev. Phys. Chem.* **2014**, *65*, 149–174.
- (4) Nerenberg, P. S.; Head-Gordon, T. Optimizing Protein–Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011**, *7* (4), 1220–1230.
- (5) Best, R. B.; Mittal, J. Free-Energy Landscape of the GB1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins* **2011**, *79* (4), 1318–1328.
- (6) Best, R. B.; Mittal, J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010**, *114* (46), 14916–14923.
- (7) Adhikari, A. N.; Freed, K. F.; Sosnick, T. R. Simplified Protein Models: Predicting Folding Pathways and Structure Using Amino Acid Sequences. *Phys. Rev. Lett.* **2013**, *111* (2), 028103.
- (8) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.
- (9) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. Structure and Dynamics of the $A\beta_{21-30}$ Peptide from the Interplay of NMR Experiments and Molecular Simulations. *J. Am. Chem. Soc.* **2008**, *130* (19), 6145–6158.
- (10) Knott, M.; Best, R. B. A Prefolded Binding Interface in the Unbound Ensemble of an Intrinsically Disordered Protein: Evidence from Molecular Simulations. *PLoS Comput. Biol.* **2012**, *8* (7), No. e1002605.
- (11) Nettels, D.; Müller-Späh, S.; Küster, F.; Hofmann, H.; Haenni, D.; Rügger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; et al. Single-Molecule Spectroscopy of the Temperature-Induced Collapse of Unfolded Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (49), 20740–20745.
- (12) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. Characterizing the Unfolded States of Proteins Using Single-Molecule FRET Spectroscopy and Molecular Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (5), 1528–1533.
- (13) Voelz, V. A.; Jäger, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; et al. Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment. *J. Am. Chem. Soc.* **2012**, *134* (30), 12565–12577.
- (14) Sosnick, T. R.; Barrick, D. The Folding of Single Domain Proteins—Have We Reached a Consensus? *Curr. Opin. Struct. Biol.* **2011**, *21* (1), 12–24.
- (15) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. Evaluating the Effects of Cutoffs and Treatment of Long-Range Electrostatics in Protein Folding Simulations. *PLoS One* **2012**, *7* (6), e39918.
- (16) Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S. Unfolded-State Dynamics and Structure of Protein L Characterized by Simulation and Experiment. *J. Am. Chem. Soc.* **2010**, *132* (13), 4702–4709.
- (17) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. Temperature-Dependent Solvation Modulates the Dimensions of Disordered Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (14), 5213–5218.
- (18) Mittag, T.; Forman-Kay, J. D. Atomic-Level Characterization of Disordered Protein Ensembles. *Curr. Opin. Struct. Biol.* **2007**, *17* (1), 3–14.
- (19) Vendruscolo, M. Determination of Conformationally Heterogeneous States of Proteins. *Curr. Opin. Struct. Biol.* **2007**, *17* (1), 15–20.
- (20) Sun, Y.; Kollman, P. A. Hydrophobic Solvation of Methane and Nonbond Parameters of the TIP3P Water Model. *J. Comput. Chem.* **1995**, *16*, 1164–1169.
- (21) 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* (15), 4524–4534.
- (22) Kolár, M.; Berka, K.; Jurecka, P.; Hobza, P. On the Reliability of the Amber Force Field and Its Empirical Dispersion Contribution for the Description of Noncovalent Complexes. *ChemPhysChem* **2010**, *11* (11), 2399–2408.
- (23) Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. Solution Structure of the N-terminal Zinc Binding Domain of HIV-1 Integrase. *Nat. Struct. Biol.* **1997**, *4* (7), 567–577.
- (24) Wikström, M.; Drakenberg, T.; Forsén, S.; Sjöbring, U.; Björck, L. Three-Dimensional Solution Structure of an Immunoglobulin Light Chain-Binding Domain of Protein L. Comparison with the IgG-Binding Domains of Protein G. *Biochemistry* **1994**, *33* (47), 14011–14017.
- (25) Kremer, W.; Schuler, B.; Harrieder, S.; Geyer, M.; Gronwald, W.; Welker, C.; Jaenicke, R.; Kalbitzer, H. R. Solution NMR Structure of the Cold-Shock Protein from the Hyperthermophilic Bacterium *Thermotoga Maritima*. *Eur. J. Biochem.* **2001**, *268* (9), 2527–2539.
- (26) Ulmer, T. S.; Bax, A.; Cole, N. B.; Nussbaum, R. L. Structure and Dynamics of Micelle-Bound Human α -Synuclein. *J. Biol. Chem.* **2005**, *280* (10), 9595–9603.
- (27) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; 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* (18), 3586–3616.
- (28) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossváry, I.; Moraes, M. A.; Sacerdoti, F. D.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)* 2006, IEEE, New York.
- (29) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511–519.
- (30) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.
- (31) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177–4189.
- (32) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; et al. Millisecond-Scale Molecular Dynamics Simulations on Anton. *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)* 2009, ACM, New York.
- (33) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990–2001.

- (34) Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comput. Chem.* **2001**, *22* (5), 501–508.
- (35) Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Gaussian Split Ewald: A Fast Ewald Mesh Method for Molecular Simulation. *J. Chem. Phys.* **2005**, *122* (5), 54101.
- (36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *Amber 12*; University of California: San Francisco, CA, 2012.
- (37) 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* **2010**, *78* (8), 1950–1958.
- (38) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725.
- (39) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113* (26), 9004–9015.
- (40) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100* (9), L47–L49.
- (41) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105* (28), 6474–6487.
- (42) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. Force Field Validation Using Protein Side Chain Prediction. *J. Phys. Chem. B* **2002**, *106* (44), 11673–11680.
- (43) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids: 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103* (2), 335–340.
- (44) Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. J. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665–9678.
- (45) Abascal, J. L. F.; Vega, C. J. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 234505.
- (46) Amber12 and Amber99SB-ILDN are very similar force fields, but Amber12 tends to yield unfolded states characterized by more helical content than Amber99SB-ILDN. A comparison between the two force fields thus allows one to evaluate the effect of a force field's propensity for the formation of a particular type of secondary structure. We chose to focus further on the Amber suite of force fields because, in principle, they have not been parametrized to be used with a particular water model; they are thus expected to be slightly more transferable across different water models.
- (47) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2012**, *134* (8), 3787–3791.
- (48) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91* (1), 461–466.
- (49) Casimir, H. B. G.; Polder, D. The Influence of Retardation on the London-van der Waals Forces. *Phys. Rev.* **1948**, *73*, 360–372.
- (50) Donchev, A. G.; Galkin, N. G.; Tarasov, V. I. Anisotropic Nonadditive Ab Initio Force Field for Noncovalent Interactions of H₂. *J. Chem. Phys.* **2007**, *126* (17), 174307.
- (51) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94* (7), 1887–1930.
- (52) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: A General-Purpose Quantum Chemistry Program Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 242–253.
- (53) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; et al. MOLPRO, version 2012.1, a Package of Ab Initio Programs; see <http://www.molpro.net/>.
- (54) 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* (12), 10089–10092.
- (55) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
- (56) Vega, C.; de Miguel, E. Surface Tension of the Most Popular Models of Water by Using the Test-Area Simulation Method. *J. Chem. Phys.* **2007**, *126* (15), 154707.
- (57) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (58) Zheng, R.; Jenkins, T. M.; Craigie, R. Zinc Folds the N-Terminal Domain of HIV-1 Integrase, Promotes Multimerization, and Enhances Catalytic Activity. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (24), 13659–13664.
- (59) We focus most of our comparison on R_g because it is a global quantity that can be estimated with different experimental methods, and because we generally observe large qualitative discrepancies between experimental estimates and computational results. Also, a rough initial estimate of R_g can be obtained from simulations only a few microseconds long (Figure S1, Supporting Information), whereas converging other structural properties often requires much longer time scales.
- (60) Müller-Späh, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rügger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge Interactions Can Dominate the Dimensions of Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (33), 14609–14614.
- (61) Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O'Hern, C. S.; Rhoades, E. The Conformational Ensembles of α -Synuclein and Tau: Combining Single-Molecule FRET and Simulations. *Biophys. J.* **2012**, *103* (9), 1940–1949.
- (62) Note that recent TIP4P-like models, parametrized based on a much more extended data set, feature dispersion coefficients larger than earlier models.^{45,88}
- (63) Rijks, W.; Wormer, P. E. S. Correlated van der Waals Coefficients: II. Dimers Consisting of CO, HF, H₂O, and NH₃. *J. Chem. Phys.* **1989**, *90* (11), 6507–6519.
- (64) Wormer, P. E. S.; Hetteema, H. Many-Body Perturbation Theory of Frequency-Dependent Polarizabilities and van der Waals Coefficients: Application to H₂O–H₂O and Ar–NH₃. *J. Chem. Phys.* **1992**, *97* (8), 5592–5606.
- (65) London, F. The General Theory of Molecular Forces. *Trans. Faraday Soc.* **1937**, *33*, 8–26.
- (66) A small dispersion coefficient may also be in part the result of overfitting to the density and position of the first peak of the RDF, which are strongly influenced by the repulsive C_{12}/r^{12} term. Unfortunately, this computationally convenient functional form provides overly strong repulsion.^{89–91} In order to obtain a reasonable description of the first peak of the RDF using this steep repulsive functional form, a weak attractive dispersion coefficient must be used. (See the Supporting Information for a more in-depth discussion.)
- (67) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**, *117* (34), 9956–9972.
- (68) Halgren, T. A. Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *J. Am. Chem. Soc.* **1992**, *114* (20), 7827–7843.
- (69) Gubskaya, A.; Kusalik, P. G. The Total Molecular Dipole Moment for Liquid Water. *J. Chem. Phys.* **2002**, *117* (11), 5290–5302.
- (70) Silvestrelli, P. L.; Parrinello, M. Water Molecule Dipole in the Gas and in the Liquid Phase. *Phys. Rev. Lett.* **1999**, *82* (16), 3308–3311.

(71) Note that nonpolarizable water models are not expected to reproduce the liquid–vapor equilibrium with high accuracy, as they cannot correctly describe the water dipole moment in both the liquid and gas phases.

(72) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103* (22), 4570–4577.

(73) Morar, A. S.; Olteanu, A.; Young, G. B.; Pielak, G. J. Solvent-Induced Collapse of α -Synuclein and Acid-Denatured Cytochrome *c*. *Protein Sci.* **2001**, *10* (11), 2195–2199.

(74) Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E.; et al. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α -Synuclein in Solution from NMR and Small Angle Scattering. *Structure* **2014**, *22* (2), 238–249.

(75) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105* (4), 962–974.

(76) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. FoXS: A Web Server for Rapid Computation and Fitting of SAXS Profiles. *Nucleic Acids Res.* **2010**, *38*, W540–544.

(77) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in α -Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127* (2), 476–477.

(78) Bertocini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. Release of Long-Range Tertiary Interactions Potentiates Aggregation of Natively Unstructured α -Synuclein. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (5), 1430–1435.

(79) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132* (24), 8407–8418.

(80) Sung, Y. H.; Eliezer, D. Residual Structure, Backbone Dynamics, and Interactions within the Synuclein Family. *J. Mol. Biol.* **2007**, *372* (3), 689–707.

(81) Esteban-Martín, S.; Silvestre-Ryan, J.; Bertocini, C. W.; Salvatella, X. Identification of Fibril-Like Tertiary Contacts in Soluble Monomeric α -Synuclein. *Biophys. J.* **2013**, *105* (5), 1192–1198.

(82) Ozenne, V.; Schneider, R.; Yao, M.; Huang, J. R.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M. Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. *J. Am. Chem. Soc.* **2012**, *134* (36), 15138–15148.

(83) Robustelli, P.; Trbovic, N.; Friesner, R. A.; Palmer, A. G., III. Conformational Dynamics of the Partially Disordered Yeast Transcription Factor GCN4. *J. Chem. Theory Comput.* **2013**, *9* (11), 5190–5200.

(84) Shen, Y.; Bax, A. SPARTA+: A Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol.* **2010**, *48* (1), 13–22 NMR.

(85) De Simone, A.; Cavalli, A.; Hsu, S. T.; Vranken, W.; Vendruscolo, M. Accurate Random Coil Chemical Shifts from an Analysis of Loop Regions in Native States of Proteins. *J. Am. Chem. Soc.* **2009**, *131* (45), 16332–16333.

(86) Lee, J. C.; Gray, H. B.; Winkler, J. R. Tertiary Contact Formation in α -Synuclein Probed by Electron Transfer. *J. Am. Chem. Soc.* **2005**, *127* (47), 16388–16389.

(87) Grupi, A.; Haas, E. J. Segmental Conformational Disorder and Dynamics in the Intrinsically Disordered Protein α -Synuclein and Its Chain Length Dependence. *J. Mol. Biol.* **2011**, *405* (5), 1267–1283.

(88) Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A Potential Model for the Study of Ices and Amorphous Water: TIP4p/Ice. *J. Chem. Phys.* **2005**, *122* (23), 234511.

(89) Warshel, A.; Shneior, L. Consistent Force Field Calculations. II: Crystal Structures, Sublimation Energies, Molecular and Lattice Vibrations, Molecular Conformations, and Enthalpies of Alkanes. *J. Chem. Phys.* **1970**, *53* (2), 582–594.

(90) Kramer, C.; Gedeck, P.; Meuwly, M. Multipole-Based Force Fields from Ab Initio Interaction Energies and the Need for Jointly

Refitting All Intermolecular Parameters. *J. Chem. Theory Comput.* **2013**, *9* (3), 1499–1511.

(91) Sun, H. COMPASS: An Ab Initio Force-Field Optimized for Condensed-Phase Applications—Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102* (38), 7338–7364.

(92) Sherman, E.; Haran, G. Coil–Globule Transition in the Denatured State of a Small Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (31), 11539–11543.

(93) Uversky, V. N.; Gillespie, J. R.; Millett, I. S.; Khodyakova, A. V.; Vasilenko, R. N.; Vasiliev, A. M.; Rodionov, I. L.; Kozlovskaya, G. D.; Dolgikh, D. A.; Fink, A. L.; et al. Zn²⁺-Mediated Structure Formation and Compaction of the “Natively Unfolded” Hyman Prothymosin α . *Biochem. Biophys. Res. Commun.* **2000**, *267*, 663–668.

(94) Zweckstetter, M.; Bax, A. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **2000**, *122* (15), 3791–3792.

(95) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10* (11), 5113–5124.

■ NOTE ADDED IN PROOF

After submission of this work, Best and coworkers⁹⁵ have shown that, as suggested by previous work,²¹ optimization of the solute–solvent dispersion interaction against experimental data can also result in more expanded disordered protein ensembles that agree well with experiment. The expansion seems to be driven by the increase in solute–solvent interactions, as reflected by the differences in the hydration free energies of small molecules. There are clearly similarities with the present study, as increasing the water dispersion coefficient also results in increased solute–solvent dispersion interactions. However, in this case, additional changes are also introduced that have the net effect of leaving the hydration free energy of small molecules almost unchanged. This observation raises the intriguing question of whether these studies, while sharing several similarities, may be describing a different underlying physics.