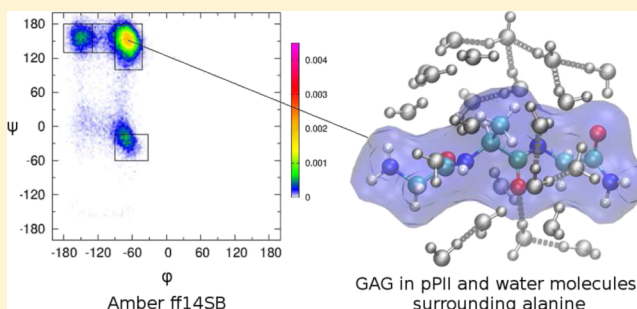


Do Molecular Dynamics Force Fields Capture Conformational Dynamics of Alanine in Water?

Shuting Zhang,[†] Reinhard Schweitzer-Stenner,^{‡,§} and Brigita Urbanc^{*,†,§}[†]Department of Physics and [‡]Department of Chemistry, Drexel University, Philadelphia, Pennsylvania 19104, United States

Supporting Information

ABSTRACT: We examine the ability of six molecular dynamics (MD) force fields (Amber ff14SB, Amber ff99SBnmr1, Amber ff03ws, OPLS-AA/L, OPLS-AA/M, and CHARMM36) to reproduce conformational ensembles of the central alanine in GAG and AAA in a way that is consistent with five (GAG) or six (AAA) *J* coupling constants and amide I' profiles. MD-derived Ramachandran plots for all six force fields under study differ from those obtained by the Gaussian fit to experimental data in three major ways: (i) the polyproline II (pPII) basin in the Ramachandran plot is too concentrated, (ii) the antiparallel β ($\alpha\beta$) basin is overpopulated, and (iii) the transitional β (β_t) basin is underpopulated. Amber ff14SB outperforms the other five MD force fields and yields the highest pPII populations of the central alanine residue in GAG (55%) and AAA (63%), in good agreement with the predictions of the Gaussian model (59 and 76%). The analysis of the hydration layer around the central alanine residue reveals considerable reorientation of water molecules and reduction in both the average number of water molecules and the average number of water–water hydrogen bonds when glycines (in GAG) are replaced by alanines (in AAA), elucidating water-mediated nearest neighbor effects on alanine's conformational dynamics.



INTRODUCTION

Intrinsically disordered proteins (IDPs) defy the classical view of the protein function arising from a stable three-dimensional structure.¹ More than 30% of gene sequences in eukaryotic genomes encode proteins or protein regions that lack a well-structured three-dimensional fold, yet these unstructured regions are often key to the function of IDPs.^{2,3} Structural characterization of IDPs is challenging both experimentally and computationally. Molecular dynamics (MD) simulations can provide key insights into the structural dynamics of IDPs and, by implication, also their function. However, reliability of structural predictions derived from MD simulations depends on the accuracy of the underlying force field. Despite formidable progress in the development of accurate all-atom MD force fields in the past decade,⁴ MD force fields fail to accurately reproduce conformational dynamics of intrinsic amino acid residues and unfolded short peptides, which are the building blocks of the unstructured regions of IDPs. Specifically, the affinity of most amino acid residues (in particular alanine) to adopt the polyproline II (pPII) conformation is typically underestimated, whereas the helical content is overestimated.^{5–7} We here posit that a successful force field for IDPs alongside the appropriate water model should be able to reproduce conformational dynamics of short unfolded peptides and proteins in agreement with the experimental data.

To assess intrinsic conformational propensities of amino acid residues in water, blocked dipeptides and glycine-based host–guest systems are typically used in computational and spectroscopic studies because the terminal charges of unblocked (tri)peptides are generally suspected to exert undesired effects on conformational propensities of amino acid residues.^{8,9} However, this notion is at variance with the results of Toal et al. who showed that the conformational ensembles of the central alanine in AAA are not affected by the protonation states of the terminal groups.⁶ Moreover, they found that alanine in AdP and GAG displayed very similar conformational propensities, suggesting that dipeptides and unblocked GxG peptides are both suitable systems for studying intrinsic conformational propensities of amino acid residues in water.

The intrinsic conformational ensembles (Ramachandran distributions) of amino acid residues in water have been determined by measuring and analyzing several scalar NMR coupling constants and amide I' band profiles in the IR, polarized Raman, and vibrational circular dichroism (VCD) spectra in cationic GxG tripeptides.^{10–15} The term “intrinsic” indicates that the influence of the adjacent residues (i.e., glycine) on the conformational distribution of the guest residue *x* is assumed to be negligible. To relate the scalar NMR

Received: June 13, 2019

Published: November 21, 2019

coupling constants and amide I' profiles to conformational distributions in the Ramachandran space, Schweitzer-Stenner developed a Gaussian model, in which Ramachandran distributions for the central amino acid residues of tripeptides are described as a superposition of two-dimensional Gaussian subdistributions related to different conformational states, such as pPII, β -strand, right- and left-handed helical, and various turn-like conformations.¹⁰ The positions, half-widths, and statistical weights of Gaussian model subdistributions are adjusted to best fit the average J coupling constants, which are calculated using Karplus equations (Graf et al.¹⁶ and references cited therein) and amide I' profiles, for which the excitonic coupling model, describing the intensity distributions of amide I' profiles in terms of conformationally dependent coupling between amide I' modes of different peptide groups, is used.¹⁷ This theoretical approach was used to obtain Ramachandran plots and conformational propensities of 16 unblocked GxG peptides in water,^{12–14,18} revealing the prevalence of the pPII state in most amino acid residues and the resulting conformational dynamics dominated by pPII/ β -strand equilibria. To what extent the current MD force fields capture the experimentally observed intrinsic conformational dynamics of amino acid residues in water remains an open question.

MD studies of peptides and proteins are dominated by three classes of all-atom MD force fields for biomolecular simulations: Optimized Potentials for Liquid Simulations (OPLS), Chemistry at HARvard Macromolecular Mechanics (CHARMM), and Assisted Model Building with Energy Refinement (Amber) combined with various water models, such as TIP32, TIP3P, and TIP4P;¹⁹ SPC/E;²⁰ and the more recent TIP4P/2005 water model, which reproduces the water density maximum at 4 °C as well as a number of other anomalous water properties.²¹ Whereas the parameterization of the all-atom OPLS (OPLS-AA) force field was initially based on the experimental data,^{22,23} torsional coefficients were later modified based on quantum mechanical (QM) calculations, resulting in the OPLS-AA/L force field.²⁴ Recently, Jorgensen and collaborators further optimized this force field with respect to backbone dihedral and side-chain torsional potentials to fit ab initio gas-phase QM calculations of alanine and glycine dipeptides alongside side-chain torsional potential scans for the remaining dipeptides, resulting in OPLS-AA/M.²⁵ Similar empirical potentials were used in the CHARMM development, whereby CHARMM22 was parametrized alongside the TIP3P water model.²⁶ An improved version of CHARMM22, CHARMM22-CMAP (also known as CHARMM27), introduced an extension of the potential energy function to include backbone dihedral angle crossterms (CMAP correction) but also pointed out limitations of gas-phase QM calculations in reproducing the experimental crystallographic data.²⁷ Best et al. optimized the backbone dihedral and side-chain torsional potential functions against solution NMR data and NMR scalar couplings derived for weakly structured and unfolded peptides, which reduced the α -helical bias of the CHARMM22/CMAP version, resulting in CHARMM36.²⁸ The Amber force field, ff94, for protein simulations with the TIP3P water model was released in 1995,²⁹ followed by a number of Amber variants, including ff99,³⁰ ff03,³¹ and notably, ff99SB,³² which is the “canonical” Amber force field, commonly used as a starting force field for improved parameterization, resulting in several revised Amber force fields released over the recent years.^{33–39} The canonical Amber ff99SB improved dihedral potentials in the ff99 energy

function based on energies obtained by gas-phase ab initio QM calculations on glycine and alanine tetrapeptides, resulting in a better secondary structure balance for glycine and alanine with respect to PDB⁴⁰ survey data.³² Li and Brüschweiler employed NMR chemical shifts of four trial proteins, two with α -helical and two with $\alpha + \beta$ folds, to optimize the backbone dihedral potentials of ff99SB for simulating proteins with the SPC/E (instead of the more common TIP3P) water model, resulting in the ff99SB-nmr1 force field.³⁴ Best et al. used Amber ff03³¹ with TIP4P/2005 and modified the backbone dihedral parameters, resulting in Amber ff03w, which lead to increased cooperativity of the helix–coil transition and, unlike its predecessor ff03, captured the experimentally observed temperature-induced collapse in the radius of gyration of folded peptides.⁴¹ Best et al. further modified ff03w by increasing the strength of short-range protein–water interactions to increase protein solubility, resulting in the ff03ws force field, which was reported to improve properties of disordered proteins.⁴² The canonical ff99SB force field was further revised with respect to backbone dihedral potentials and side-chain torsional potentials, resulting in ff14SB force field.³⁷ Amber ff14SB introduced empirical corrections to the backbone dihedral potentials and modified side-chain torsional potentials using gas-phase ab initio QM calculations of the complete set of amino acid dipeptides.³⁷

When examining Ramachandran distributions of backbone dihedral angles ϕ and ψ of amino acid residues, the concept of mesostate is frequently used as a population of conformations confined to a predefined rectangular region of the (ϕ, ψ) space.^{43–45} In our previous work, we explored conformational dynamics of alanine in trialanine (AAA) and alanine dipeptide (AdP) using two MD force fields, OPLS-AA/L and Amber ff03, combined with several water models and found that OPLS-AA/L combined with SPC/E best matched the experimentally derived pPII and β -strand mesostate populations.⁶ In a subsequent study, we examined conformational dynamics of 15 guest residues x in GxG peptides by MD using OPLS-AA/L combined with SPC/E and TIP3P to demonstrate that the correspondence between experimental and MD-derived mesostate populations is less than satisfactory.⁷ In the past decade, a number of MD force fields with modified dihedral and torsional potentials have been proposed, whereby modifications were based on novel QM calculations, NMR data, or both. Alanine is of particular interest because of its notoriously high propensity to adopt the pPII-type conformations,^{6,10,11,16,46–48} which is typically underestimated in MD simulations, but also because QM calculations on alanine-based short peptides are often used to calibrate the force field parameters, in particular, dihedral potentials.

In this study, we examine to which extent the intrinsic conformational dynamics of alanine in two short peptides, GAG and AAA, is captured by two OPLS force fields (OPLS-AA/L and OPLS-AA/M), CHARMM36, and four different Amber force fields (ff99SB, ff03ws, ff99SB-NMR1, and ff14SB). The MD-derived Ramachandran dihedral angle distributions are then used to calculate five (GAG) or six (AAA) J coupling constants and amide I' profiles of respective Raman, IR, and VCD spectra to facilitate a direct comparison to the previously published experimental data.^{6,11,14} We use the respective differences between measured and calculated J coupling constants and VCD amide I' profiles as criteria to assess the degree to which the MD-derived Ramachandran distributions can account for the experimental data. It is

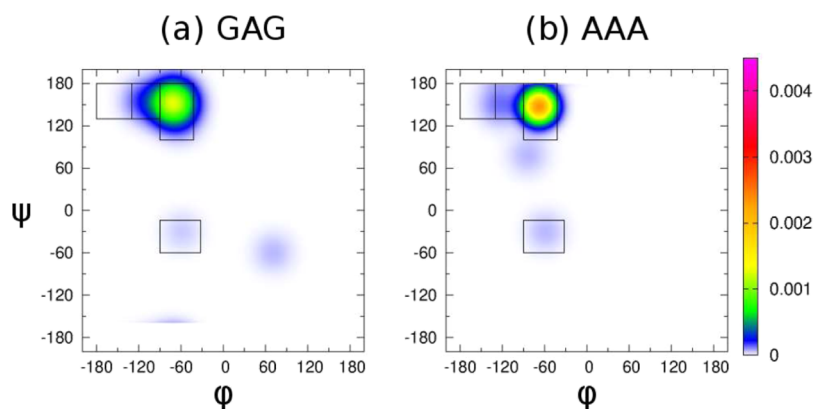


Figure 1. Ramachandran plots for alanine in GAG and AAA derived from the Gaussian model. Ramachandran plots for (a) alanine in GAG and (b) the central alanine in AAA derived by fitting the Gaussian model parameters to best reproduce experimentally obtained J coupling constants and VCD amide I' profiles.

important to note that although we describe our results in terms of mesostate populations, our assessment of the six force fields takes into consideration the entire Ramachandran distributions and is thus independent of our particular definition of the mesostates. We also study the stability of conformational populations in different force fields by calculating and comparing their lifetimes. Finally, we explore properties of hydration water around the central alanine residue in GAG and AAA, including the orientational degrees of freedom of water adjacent to the side chain, which is visualized through previously reported water orientation plots.⁷

METHODS

MD Simulations. For each force field/water model combination, dynamics of a single GAG or AAA peptide is probed at temperature 300 K using GROMACS 5.1.2.^{49–55} The peptide is capped with a protonated N-terminus ($-\text{NH}_3^+$) and neutral C-terminus ($-\text{COOH}$ for OPLS and CHARMM force fields and $-\text{CO}-\text{NH}_2$ for Amber force fields). Each peptide is solvated in a cubic box of edge length 40 Å using the corresponding water model. One Cl^- is added to neutralize the system. Cubic periodic boundary conditions and Verlet cutoff scheme⁵⁶ are applied. A time step of 2 fs is used. The energy minimization is performed using the steepest descent minimization for 100 000 steps, followed by a pressure equilibration for 20 ns at 300 K and 1.0 bar. All 100 ns (or 200 ns) long trajectories are acquired using the velocity rescale thermostat⁵⁷ and Berendsen barostat.⁵⁸

Analysis. Gaussian Model. The Gaussian model, which was introduced by Schweitzer-Stenner,¹⁰ is based on the 2D distribution function of the two backbone dihedral angles of the amino acid residue under consideration

$$P(\phi_i, \psi_j) = \sum_k \chi_k G_k(\phi_i, \psi_j) \quad (1)$$

where χ_k is the mole fraction (propensity) of the k -th state, modeled by a two-dimensional Gaussian subdistribution G_k with the maximum at $\phi_{k,\text{max}}$ and $\psi_{k,\text{max}}$ and half-widths $\sigma_{k,\phi}$ and $\sigma_{k,\psi}$. For any given dihedral angle distribution, $P(\phi_i, \psi_j)$, the J coupling constants are calculated using Karplus equations with previously reported Karplus parameters.¹⁶ Hu and Bax⁵⁹ reported two sets of Karplus parameters: one based entirely on X-ray data and another obtained from a combined analysis of solution and X-ray data. Because the latter set yields

substantially lower root mean square deviation values, it is frequently used in the literature¹⁶ and is also adopted in our present study. Figure S1 shows Karplus plots using six different Karplus parameter sets: four empirical sets reported by Bax and collaborators^{59,60} and two sets derived from DFT calculations.⁶¹ The two Karplus plots derived from DFT calculations deviate strongly from the plots produced with the four empirical sets. Importantly, Karplus plots produced with the four empirical data sets from Hu and Bax⁵⁹ and Wang and Bax⁶⁰ are very similar to each other, in particular for negative ϕ -values. Only for positive ϕ -values, noticeable differences in the $^3J(\text{H}^N, \text{C}_\beta)$ maximum are observed but these difference would be relevant only for Ramachandran distributions with a substantial population of left-handed helical conformations, which is not the case for alanine.

The corresponding amide I' band profile are derived for each of the $180 \times 180 = 32\,400$ ϕ and ψ values and averaged (the prime sign in I' indicates that the measurements were performed in D_2O). More specifically, if $f(\phi_i, \psi_i)$ describes the dependence of the amide I' profile or J coupling constant on the two dihedral angles, then the average value is calculated as

$$\langle f \rangle = \sum_{i,j} f(\phi_i, \psi_j) P(\phi_i, \psi_j) \quad (2)$$

As reported by Hagarman et al., the parameters of the Gaussian model (χ_k values, positions, and widths of the corresponding Gaussian subdistributions) are then adjusted in such a way that the calculated J coupling constants and amide I' profiles best match the corresponding experimental J coupling constants and amide I' profiles¹¹ (see Figure 1).

The experimental data for the central alanine in cationic GAG and AAA include the following J coupling constants: $^3J(\text{H}^N, \text{H}^{\text{C}_\alpha})$, $^3J(\text{H}^N, \text{C}')$, $^3J(\text{H}^N, \text{C}_\beta)$, $^3J(\text{H}^{\text{C}_\alpha}, \text{C}')$, and $^1J(\text{N}, \text{C}_\alpha)$. The additional J coupling constant, $^3J(\text{C}', \text{C}')$, which was reported for the central alanine in AAA,^{11,16} is included in the derivation of the Gaussian model parameters for AAA.

Definition of Mesostates. The following mesostate definitions are used: (a) polyproline II (pPII) ($-90^\circ < \phi < -42^\circ$, $100^\circ < \psi < 180^\circ$), (b) antiparallel β -strand ($\text{a}\beta$) ($-180^\circ < \phi < -130^\circ$, $130^\circ < \psi < 180^\circ$), (c) the transition region between $\text{a}\beta$ and pPII (bt) ($-130^\circ < \phi < -90^\circ$, $130^\circ < \psi < 180^\circ$), and (d) right-handed α -helix ($-90^\circ < \phi < -32^\circ$, $-60^\circ < \psi < 14^\circ$) (Figure 2).

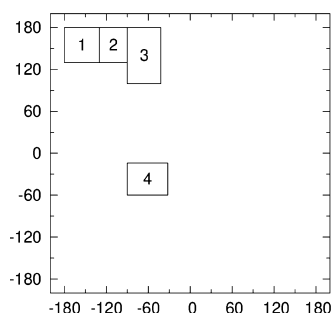


Figure 2. Definitions of the four mesostates in the Ramachandran plot. The regions 1–4 correspond to $\alpha\beta$, $\beta\beta$, pPII, and right-handed α -helix, respectively.

MD-Derived Ramachandran Plots. Ramachandran plots are constructed from dihedral angles (ϕ , ψ) within GROMACS 5.1.2 using 25 000 time frames of each MD trajectory between 50 and 100 ns (with time frames 2 ps apart). Normalized 2D distributions are calculated with a bin size of $2^\circ \times 2^\circ$, resulting in $180 \times 180 = 32\,400$ bins to facilitate a direct comparison to the Ramachandran plots predicted by the Gaussian model.

Calculation of J Coupling Constants and Amide I' Profiles from MD Data. As shown in the Results and Discussion section, nearly all MD-derived Ramachandran distributions reproduce the IR and Raman band profiles rather well because of their rather limited content of right-handed helical conformations. Therefore, the J coupling constants and the VCD band profile, which are the most sensitive to the sampling of the upper left quadrant of the Ramachandran plot, represent the best indicators of differences among different MD force fields. We use MD-derived Ramachandran distributions to calculate the J coupling constants and amide I' profiles for alanine in GAG and the central alanine in AAA. With regard to J coupling constants, we characterize the overall difference between the calculated and experimental values by the absolute value of their differences as well as by χ_J^2 function

$$\chi_J^2 = \frac{1}{N} \sum_{i=1}^N \frac{(J_{i,\text{exp}} - J_{i,\text{calc}})^2}{s_i^2} \quad (3)$$

where $J_{i,\text{exp}}$ and $J_{i,\text{calc}}$ denote experimental and calculated J coupling constants, respectively. N is the number of considered coupling constants, whereas s_i are the corresponding statistical errors. Following Maier et al., we calculate the latter as a combination of experimental errors¹¹ and errors due to the reported uncertainties of the Karplus parameters^{59,60} using Gaussian error propagation. Statistical uncertainties s_i are ensemble averages of the following function $\bar{s}_i(\phi)$

$$\begin{aligned} \bar{s}_i(\phi) &= \sqrt{s_{A_i}^2 \cos^4(\phi + \theta_i) + s_{B_i}^2 \cos^2(\phi + \theta_i) + s_{C_i}^2 + s_{J_i}^2} \end{aligned} \quad (4)$$

where s_{A_i} , s_{B_i} , and s_{C_i} are the statistical uncertainties of the three Karplus parameters listed in Table S1, θ_i is the phase associated with the J coupling constant J_i , s_{J_i} is the corresponding experimental error (as previously reported for GAG¹¹ and AAA¹⁶), and $s_i = \langle \bar{s}_i \rangle$, whereby the ensemble average $\langle \rangle$ is defined by eq 2. Wang and Bax reported s_{A_i} , s_{B_i} , and s_{C_i} values for two different sets of Karplus parameters: one obtained from

the fit of the data set derived from X-ray structures of proteins and another derived from a data set based on a combination of X-ray and NMR structures.⁶⁰ While there is no obvious reason to favor one over the other, we here selected the former set with larger s values as a conservative estimate of J -coupling uncertainties. Unfortunately, statistical errors for two Karplus parameters $^1J(\text{N}, \text{C}_\alpha)$ (relevant to GAG and AAA) and $^3J(\text{C}', \text{C}')$ (relevant only to AAA) are not known.^{16,62} For these two J coupling constants, only the experimental errors are considered in the above calculation, leading to an overestimation of χ_J^2 values. However, because this work focuses on a comparison rather than on the absolute value of χ_J^2 values, lower uncertainties for $^1J(\text{N}, \text{C}_\alpha)$ and $^3J(\text{C}', \text{C}')$ in the χ_J^2 function are mostly inconsequential.

The reproduction of the VCD band profile is assessed by

$$\chi_{\text{VCD}}^2 = \frac{1}{N'} \sum_{k=1}^{N'} \frac{(\Delta\epsilon_{\text{exp},k} - \Delta\epsilon_{\text{calc},k})^2}{s_k^2} \quad (5)$$

where $\Delta\epsilon$ represents the molar dichroism and k labels the different data point. N' is the number of data points and the standard deviation s_k is derived from an analysis of a spectral region dominated by noise.

The population of the m -th mesostate was calculated as

$$f_m = \sum_{i=\phi_{\min}^m}^{\phi_{\max}^m} \sum_{j=\psi_{\min}^m}^{\psi_{\max}^m} P(\phi_i, \psi_j) \quad (6)$$

where $P(\phi_i, \psi_j)$ is the occupation probability of dihedral angles $\phi_i \pm 1^\circ$ and $\psi_j \pm 1^\circ$, whereby the discrete values of the backbone dihedral angles are assigned in increments of 2° . The summations run from the lower bound values $\phi_{\min}^m$ and $\psi_{\min}^m$ to the upper bound values $\phi_{\max}^m$ and $\psi_{\max}^m$, respectively, which are defined for each mesostate m under consideration as described above.

Mesostate Lifetimes. Mesostate lifetimes are calculated within GROMACS 5.1.2 using all conformations between 50 and 100 ns resulting in 25 000 conformations per trajectory. Conformations are classified into mesostates following the definition above. All conformations outside the four defined mesostates are treated as a separate mesostate. The lifetime of a mesostate is defined as the time interval between alanine residue entering and exiting the mesostate. For each mesostate, a distribution of lifetimes is derived and the average lifetime τ is extracted by fitting the distribution with an exponential curve $p = p_0 e^{-t/\tau}$.

Water Orientation Plots. Water orientation plots are calculated as introduced in our previous work⁷ to elucidate the orientation of water molecules around the side chain of alanine in GAG and AAA in the pPII mesostate. Briefly, using Visual Molecular Dynamics (VMD) software package,⁶³ we extract the orientation of each water molecules in a 4 Å-thick hydration layer around side chain atoms of alanine (the methyl group $-\text{C}_\beta\text{H}_3$) relative to the local normal to the solvent accessible surface (SAS) of the side chain. The local normal to the SAS of the side chain, $\hat{n}$, is derived for each water molecule using three proximate points on the triangulated SAS area of the side chain. For each water molecule, the water symmetry axis and the normal to the water plane (i.e., the plane formed by the centers of mass of water oxygen and two water hydrogens) were determined. Each water molecule is characterized by two angles, η and θ . The angle η is defined as the angle between the water symmetry axis and $\hat{n}$. The angle

θ is computed as the angle between the normal to the water plane and $\hat{n}$. The angles η and θ of each water molecule in the hydration layer are collected into a 2D histogram using $90 \times 90 = 8100$ bins in the (η, θ) space, whereby a histogram value at (i, j) is equal to the number of water molecules $N(i, j)$ characterized by (η, θ) angles within the (i, j) -th bin. The final water orientation plot is normalized by the total number of water molecules in the hydration layer N to obtain $N(i, j)/N$, which corresponds to the probability of a water molecule to adopt (η, θ) angles within (i, j) -th bin. Water orientation plots are obtained by averaging over all pPII conformations between 50 and 100 ns of each MD trajectory.

Properties of Hydration Water around Alanine. For this analysis, which is done within GROMACS 5.1.2, we select time frames from each MD trajectory between 50 and 100 ns, which are separated by 20 ns to ensure statistical independence of consecutive configurations, then classify the configurations with respect to the mesostate of the central alanine residue in GAG and AAA. The hydration layer around alanine is chosen to include all water molecules within the center-of-mass distance of 4 Å from (1) alanine's side chain atoms ($C_\beta H_3$), (2) alanine's backbone atoms (NH, $C_\alpha H$, and CO), or the entire alanine residue (NH, $C_\alpha H$, CO, and $C_\beta H_3$). A hydrogen bond is defined by a cut-off distance of 3 Å between a donor and acceptor, using an upper limit of the hydrogen–donor–acceptor angle of 20° . The average number of water molecules in the selected hydration layer, the average number of water–water hydrogen bonds within the respective hydration layer, and the average number of alanine–water hydrogen bonds are calculated alongside their SEM values.

RESULTS AND DISCUSSION

In most MD force fields, the backbone dihedral angle potentials are parametrized by ab initio QM calculations on amino acid analogues or on glycine- and alanine-based short peptides.^{25,32,37} It is thus reasonable to expect that the intrinsic conformational propensities of alanine residues in water are well accounted for when probed by explicit-solvent MD simulations, in particular within the recently recalibrated force fields. In this paper, we assess commonly used MD force fields with respect to their ability to predict conformational ensembles of the central alanine in GAG and AAA in a way that is consistent with the experimental data. These experimental data include five (GAG) or six (AAA) NMR scalar coupling constants and amide I' profiles previously obtained by Schweitzer-Stenner and colleagues.^{6,11,14} We perform MD simulations of GAG and AAA using several MD force fields combined with their respective water model or with several commonly used water models, such as TIP3P, TIP4P, SPC/E, and TIP4P/2005. The resulting Ramachandran plots are derived by calculating probability distributions of backbone dihedral angles (ϕ, ψ) from each MD trajectory. These MD-derived probability distributions of (ϕ, ψ) values are then used to calculate the respective J coupling constants and amide I' profiles of the central alanine in GAG and AAA for a direct comparison to the experimental data as outlined in [Methods](#). To better assess the ability of a given force field to capture the experimental constraints, we compare the performance of different MD force fields to the respective performance of the Gaussian model (see [Methods](#) for more details). To facilitate a better visual comparison of Ramachandran plots produced by different MD force fields to the spectroscopic data-based Gaussian model predictions, we introduce, in the

spirit of previous studies,^{43–45} four mesostates as the most populated rectangular regions in the Ramachandran space, each defined by its mesostate-specific range of backbone dihedral ϕ and ψ angles (as defined in [Methods](#)). Assuming that a high probability density on the Ramachandran plot corresponds to a low free energy, we hereafter adhere to the free energy landscape terminology and refer to a basin as the region in the Ramachandran space (mostly coinciding with one of the four mesostates) with an increased probability density.

Conformational Ensembles of Alanine in GAG.

Gaussian Model for Alanine in GAG: Sensitivity Analysis.

Our research focuses on Ramachandran distributions of the central alanine residue in cationic GAG and AAA, whereby the performance of different MD force fields and water models is evaluated based on the ability to reproduce experimentally obtained amide I' band profiles and J -coupling constants. We utilize the previously reported Gaussian model of the Ramachandran distribution, which assumes that all conformational states can be modeled as 2D Gaussian subdistributions in the (ϕ, ψ) space (see [Methods](#) for details). Experimental data have shown that conformational dynamics of alanine in GAG and AAA is dominated by the pPII state, followed by the βt state. Because the Gaussian model is used as a benchmark for evaluating the performance of the MD force fields with respect to their ability to capture alanine conformational dynamics, we examine here the sensitivity of the Gaussian model for alanine in GAG with respect to the central (ϕ, ψ) location of the pPII state and the corresponding weight of the Gaussian pPII subdistribution. To this end, we calculated χ_j^2 , χ_{VCD}^2 , and their sum, for each of three different sets of pPII/ βt weights (0.60/0.30, 0.72/0.18, and 0.80/0.10) while keeping the weights of the right-handed helical and inverse γ turn states constant. For each set of pPII/ βt values, we then examined the effect of the central location of the pPII state by shifting the center in $\phi \in [-80^\circ, -60^\circ]$ or $\psi \in [140^\circ, 160^\circ]$ directions. [Figure S2](#) demonstrates a strong dependence of χ_j^2 on both dihedral angles, whereas χ_{VCD}^2 depends much more on ψ than on ϕ . Both χ_j^2 and χ_{VCD}^2 assume the lowest values at similar ϕ values. With regard to the ψ dependence, the situation is more complex. Whereas χ_j^2 displays a minimum at $\psi \approx 145^\circ$ and then increases with ψ , χ_{VCD}^2 decreases with ψ . It is important to note that of the five J coupling constants, only $^1J(N, C_\alpha)$ strongly depends on ψ and contributes to the ψ dependence of χ_j^2 . The optimal ψ value was thus identified as the value which minimizes the sum $\chi_j^2 + \chi_{VCD}^2$. As shown in [Figure S2](#), the minimal values of the sum of the two χ^2 values are assumed for the pPII/ βt weights of 0.8/0.1, respectively, and $(\phi = -72^\circ, \psi = 152^\circ)$. We then examined the dependence of χ_j^2 and χ_{VCD}^2 values on the location of βt in the dihedral angle space (data not shown). In view of a rather lower propensity of βt relative to pPII, it is not surprising that both χ^2 functions displayed shallow minima around their optimal positions (data not shown). Hence, the location of βt carries a rather large uncertainty of at least $\pm 15^\circ$ for ϕ and $\pm 10^\circ$ for ψ . The final Gaussian parameter values used for constructing the Ramachandran distributions of GAG and AAA within the Gaussian model are listed in [Table 1](#).

Two additional low populations that fall out of the Ramachandran space region defined by the four mesostates, that is, the inverse γ turn (GAG) and γ turn (AAA), were included in the Gaussian models as a compromise in the optimization of χ_j^2 and χ_{VCD}^2 functions. By eliminating the inverse γ turn population from the Gaussian model of GAG,

Table 1. Gaussian Model Parameters for the Central Alanine in GAG and AAA^a

pPII	GAG	AAA	βt	GAG	AAA
ϕ_{pPII}	-72°	-69°	$\phi_{\beta t}$	-115°	-125°
ψ_{pPII}	152°	147°	$\psi_{\beta t}$	155°	150°
A_{pPII}	0.80	0.84	$A_{\beta t}$	0.10	0.08

^aThe central locations of the Gaussian subdistributions corresponding to the pPII and βt states are specified as $(\phi_{\text{pPII}}, \psi_{\text{pPII}})$ and $(\phi_{\beta t}, \psi_{\beta t})$, respectively. The corresponding weights are denoted as A_{pPII} and $A_{\beta t}$, respectively.

the χ_{VCD}^2 decreased to 2.07; however, the χ_f^2 value increased to 4.9. Similarly, by eliminating the γ turn population from the Gaussian model of AAA, χ_{VCD}^2 decreased to 2.85, whereas the χ_f^2 value increased to 5.4. These relatively minor populations were included in the Gaussian models of GAG and AAA to minimize χ_f^2 while keeping χ_{VCD}^2 at reasonably low values.

Selection of the Water Model for OPLS-AA/L and OPLS-AA/M. Six force fields from three different series (OPLS, Amber, and CHARMM): OPLS-AA/L,²⁴ OPLS-AA/M,²⁵ ff99SBnmr1,³⁴ ff03ws,⁴² ff14SB,³⁷ and CHARMM36²⁸ are selected for the assessment of their capability to predict conformational distributions of alanine in GAG. Whereas Amber force fields and CHARMM36 were parametrized each with the specific water model, no specific water model was used in the parameterization of OPLS-AA/L and OPLS-AA/M. For this reason, we first examine to which extent the water model affects the conformational ensembles of alanine in GAG within OPLS-AA/L and OPLS-AA/M by comparing the Ramachandran plots derived from MD simulations using four different water models: SPC/E,¹⁹ TIP3P,⁶⁴ TIP4P,⁶⁴ and TIP4P/2005.²¹ The effect of the water model on the Ramachandran plots of alanine in GAG derived in OPLS-AA/L and OPLS-AA/M is displayed in Figures S3 and S5.

Ramachandran plots for OPLS-AA/L (Figure S3) show that regardless of the water model, pPII and α -helical conformations are overall shifted to smaller ϕ values (for pPII) and smaller ϕ and larger ψ values (for α helix), whereas the $\alpha\beta$ conformations fit well with our definition of this mesostates (Figure 1). The α -helical population appears to be favored by TIP3P in both OPLS-AA force fields (Figures S3 and S5).

The comparison of J coupling constants and amide I' profiles derived from MD simulations with OPLS-AA/L to the experimental data in Figure S4 reveals major discrepancies between experimental and calculated J coupling constants. Overall, χ_f^2 values for all OPLS-AA simulations are more than an order of magnitude larger (in the 20–40 region) than the value obtained for the Gaussian model (0.81). A similar conclusion can be made for the χ_{VCD}^2 values derived for both OPLS-AA force fields. The large χ_{VCD}^2 values reflect a substantial underestimation of the amide I' mode's rotational strength. Ramachandran plots for OPLS-AA/M in Figure S5 show that pPII conformations are better overlapped with our definition of this mesostate than in the case of OPLS-AA/L; however, the shift of α -helical conformations is comparable to the shift observed for OPLS-AA/L. OPLS-AA/M results in a notable decrease in α -helical mesostate populations relative to OPLS-AA/L. Importantly, alanine conformations outside the four mesostate regions, which are present in Ramachandran plots for OPLS-AA/L (Figure S3), are absent from the Ramachandran plots obtained with OPLS-AA/M (Figure S5). The comparison to the experimental data in Figure S4 demonstrates a significant improvement of OPLS-AA/M over OPLS-AA/L with respect to two J coupling constants ($^3J(\text{H}^{\text{C}}, \text{C}')$ and $^1J(\text{N}, \text{C}_\alpha)$) except when OPLS-AA/M is combined with TIP3P. Overall, TIP4P combined with OPLS-AA/M yields a better agreement with the experimental data than TIP3P (Figure S6). Consequently, we select TIP4P

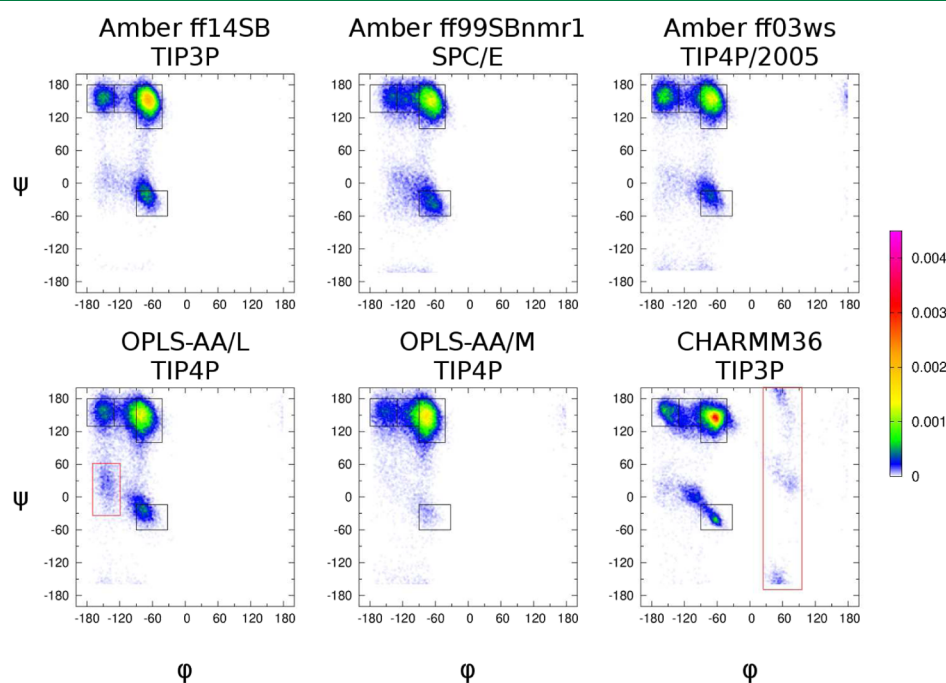


Figure 3. Ramachandran distributions for alanine in GAG. Normalized distributions of backbone dihedral angles ϕ and ψ are derived from MD simulations with different force fields, each combined with the respective water model. Black frames correspond to the four mesostates ($\alpha\beta$, βt , pPII, and α -helix) defined in Methods. The color of each $2^\circ \times 2^\circ$ bin corresponds to the occupation probability as marked on the color scale.

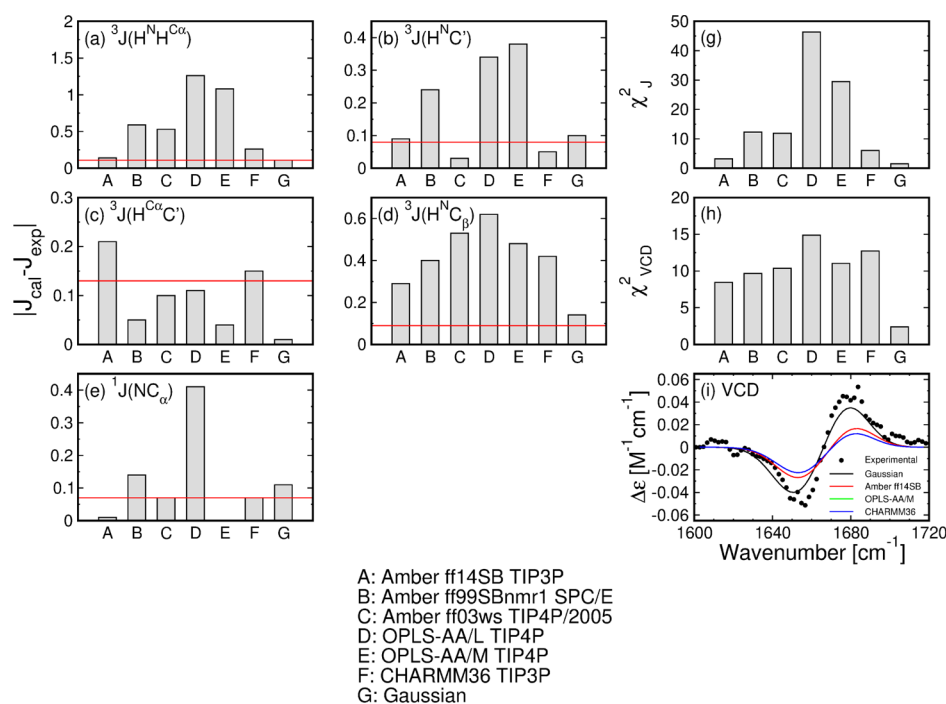


Figure 4. Comparison between MD-derived and experimental J coupling constants and VCD amide I' profiles for alanine in GAG. Five J coupling constants and VCD amide I' profile are calculated from the Ramachandran distribution obtained for each force field as well as the Gaussian model to facilitate a direct comparison to the experimental data. (a–e) Absolute differences between the calculated and experimental values ($|J_{\text{cal}} - J_{\text{exp}}|$) are displayed for each of the five J coupling constants. The red lines correspond to experimental errors. (g,h) χ^2_J and χ^2_{VCD} values as defined in [Methods](#) measure overall deviations from experimental data. (i) Calculated VCD profiles predicted by the Gaussian model, Amber ff14SB (with TIP3P), OPLS-AA/M (with TIP4P), and CHARMM36 (with TIP3P) are compared to the experimental VCD profile.

for the follow-up comparison of both OPLS-AA force fields to other force field/water model combinations.

Force Field-Specific Ramachandran Plots for Alanine in GAG. Six force fields are selected for the assessment of conformational dynamics of alanine in GAG: ff14SB (with TIP3P), ff99SBnmr1 (with SPC/E), ff03ws (with TIP4P/2005), OPLS-AA/L and OPLS-AA/M each combined with TIP4P, and CHARMM36 (with TIP3P). To probe the convergence of MD simulations, MD trajectories with Amber ff14SB, OPLS-AA/M, and CHARMM36 are extended to 200 ns and Ramachandran distributions are calculated in time windows of 50 ns ([Figure S7](#)). The comparison of Ramachandran plots across four different time windows demonstrates that despite small fluctuations in the mesostate populations, the two-dimensional dihedral angle distributions remain self-similar within each force field, indicating that 100 ns is long enough to capture the conformational dynamics of alanine in GAG. The corresponding J coupling constants calculated for 50 ns-long time intervals along each 200 ns-long trajectory also indicate that the fluctuations in J coupling constants are not large enough to affect our assessment of force fields ([Figure S8](#)). Extracting alanine conformations from the 200 ns-long trajectory of GAG in water (with time frames 2 ps apart) within each of the three force field/water model combinations, we use consecutive 10 ns-long time intervals each containing 5000 conformations to calculate the average mesostate populations. The results in [Figure S9](#) show fluctuations in the four different mesostate populations for each of the three MD force fields. Fluctuations in pII and α -helical populations are the highest in CHARMM36 and the lowest in OPLS-AA/M. The above analysis suggests that the average mesostate populations alone do not provide sufficient

information about conformational dynamics of short peptides in water and that other characteristics of the Ramachandran distribution, such as the centers and widths of mesostate basins and dynamic properties as captured by mesostate fluctuations, may be required for a more complete assessment of each force fields.

Although clearly dominated by pII populations, the six Ramachandran plots for alanine in GAG in [Figure 3](#) show distinct features with respect to the four mesostate populations, central locations of the mesostate basins, and their widths. All MD-derived Ramachandran distributions are quite distinct from the Gaussian model distribution ([Figure 1a](#)), which is characterized by a visibly broader pII basin than any MD-derived distribution. MD-derived distributions also display a well-defined $\alpha\beta$ mesostate basin that is clearly distinct from the pII basin and significantly more populated than the corresponding basin in the Gaussian model distribution, where the pII basin strongly overlaps with $\alpha\beta$ and βt basins. Most force fields produce α helical populations that are partially outside the α -helical mesostate, shifted to smaller ϕ and larger ψ values. Of the six force fields, OPLS-AA/L produces the most conformations on the left side of the Ramachandran plot that are outside the four mesostate regions ([Figure 3](#), red frame in the bottom leftmost Ramachandran plot). CHARMM36 is the only force field that gives rise to nonzero populations on the right side of the Ramachandran plot ([Figure 3](#), red frame in the bottom rightmost Ramachandran plot). In both OPLS-AA force fields, the center of the pII basin is shifted to more negative ϕ values relative the other force fields and relative to our definition of the pII mesostate. OPLS-AA/M results in the lowest α -helical population, whereas Amber ff14SB displays the lowest βt

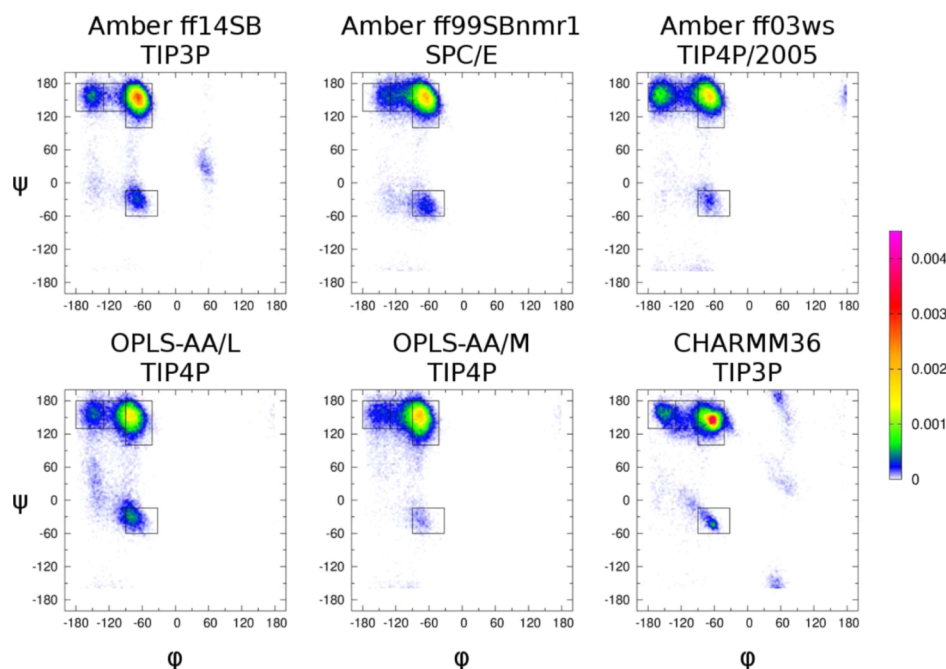


Figure 5. Ramachandran distributions for the central alanine in AAA. Normalized distributions of backbone dihedral angles ϕ and ψ are derived from MD simulations with different force fields, each combined with the respective water model. Black frames correspond to the four mesostates ($\alpha\beta$, βt , pPII, and α -helix) defined in [Methods](#). The color of each $2^\circ \times 2^\circ$ bin corresponds to the occupation probability as marked on the color scale.

population. Interestingly, for Amber ff99SBnmr1, the $\alpha\beta$ basin is shifted to larger ϕ values, that is, closer to the pPII mesostate, which is in better agreement with the Gaussian model predictions ([Figure 1a](#)).

Experimental Data for Alanine in GAG Are Best Reproduced by Amber ff14SB. A comparison of the Ramachandran distributions for alanine in GAG obtained with six different force fields ([Figure 3](#)) to the corresponding Gaussian model ([Figure 1a](#)) reveals several similarities but also elucidates major differences. Similarities include the predominant pPII population, which is in a stark contrast to the classical plots for the alanine dipeptide first reported by Ramachandran and Flory,^{65,66} and later reproduced MD force fields, such as CHARMM22⁶⁷ and the force fields included in this study. While the MD-derived Ramachandran distributions in [Figure 3](#) look qualitatively similar, they clearly differ in their capabilities to reproduce the experimental data, as shown in [Figure 4](#). The J coupling constants calculated from simulations using different force fields and Gaussian model alongside experimental values and uncertainties used in χ_f^2 calculation are reported in [Table S2](#). Considering an overall performance with respect to J coupling constants as assessed by χ_f^2 ([Figure 4g](#)), only Amber ff14SB (with TIP3P) produces a value of 3.18, which is relatively close to the prediction of the Gaussian model of 1.49. CHARMM36 (with TIP3P) ranks second with a χ_f^2 value of 6.02. The other force fields, in particular OPLS-AA/L and OPLS-AA/M, result in much higher χ_f^2 values. A comparison of the absolute differences between MD-derived and experimental J coupling constants offers important insights. All force fields under study other than Amber ff14SB significantly overestimate $^3J(\text{H}^{\text{N}}, \text{H}^{\text{C}\alpha})$. This reflects a shift of the pPII basin in the Ramachandran space toward more negative ϕ values relative to its location predicted by the Gaussian model ([Figure 1a](#)). These shifts are also responsible for the underestimation of the amide I' VCD signal ([Figure](#)

4i). Even Amber ff14SB does not account for a sufficient rotational strength of amide I' profile to match the corresponding profile derived from the Gaussian model. The reduced VCD signals reflect the ϕ dependence of the excitonic coupling between the two amide I' modes, which decreases with increasing negative ϕ values. With the sole exception of $^1J(\text{N}, \text{C}_\alpha)$, the Gaussian model outperforms all MD force fields, but this is of little significance because with the exception of OPLS-AA/L, all force fields and the Gaussian models yield $^1J(\text{N}, \text{C}_\alpha)$ values within the statistical error of the respective experimental value. CHARMM36 yields a relatively low χ_f^2 value because it reproduces $^3J(\text{H}^{\text{H}}, \text{C}')$ and $^1J(\text{N}, \text{C}_\alpha)$ quite well, while not being too much off-target on the remaining coupling constants. Thus, of all the force fields under study, Amber ff14SB outperforms the other five force fields for 3 of the 5 J coupling constants: $^3J(\text{H}^{\text{N}}, \text{H}^{\text{C}\alpha})$, $^1J(\text{N}, \text{C}_\alpha)$, and $^3J(\text{H}^{\text{N}}, \text{H}^{\text{C}\beta})$ ([Figure 4a,d,e](#)), resulting in the best agreement with the spectroscopic data for alanine in GAG.

We then asked to what extent the MD force fields under study capture IR and Raman band profiles of amide I' band, so that these additional spectroscopic data might be used as additional benchmarks. Because all MD force fields under study predominantly sample alanine conformations in the upper left quadrant of the Ramachandran plot, they are not expected to differ in their capabilities to reproduce the IR and Raman band profiles. Indeed, all MD force fields predict almost identical isotropic Raman, anisotropic Raman, and IR amide I' profiles (data not shown). [Figure S10](#) compares both Raman and IR amide I' profiles as predicted by Amber ff14SB (with TIP3P) to the corresponding experimental profiles and those derived from the Gaussian model to demonstrate that these profiles are well-reproduced by MD.

The above results suggest that among the investigated force fields, Amber ff14SB (with TIP3P) performs the best with respect to the reproduction of the J coupling constants,

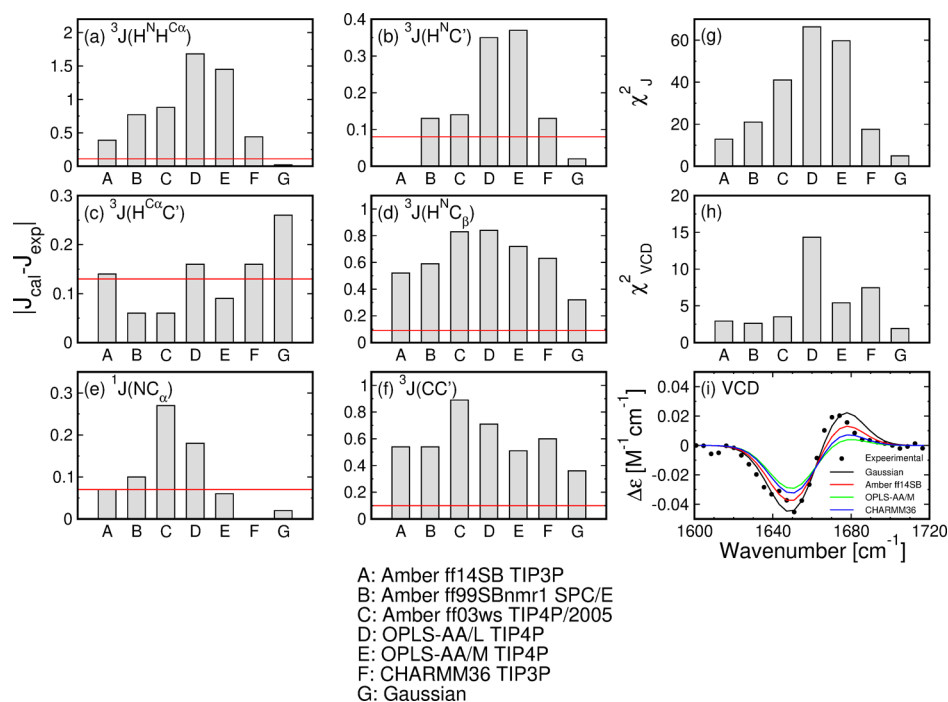


Figure 6. Comparison between MD-derived and experimental J coupling constants and VCD amide I' profiles for the central alanine in AAA. Six J coupling constants and VCD amide I' profile are calculated from the Ramachandran distribution obtained for each force field as well as the Gaussian model to facilitate a direct comparison to the experimental data. (a–f) Absolute differences between the calculated and experimental values ($|J_{\text{cal}} - J_{\text{exp}}|$) are displayed for each of the six J coupling constants. The red lines correspond to experimental errors. (g,h) χ^2_J and χ^2_{VCD} values as defined in [Methods](#) measure overall deviations from experimental data. (i) Calculated VCD profiles predicted by the Gaussian model, Amber ff14SB (with TIP3P), OPLS-AA/M (with TIP4P), and CHARMM36 (with TIP3P) are compared to the experimental VCD profile.

followed by CHARMM36 (with TIP3P), Amber ff99SBnmr1 (with SPC/E), and ff03ws (with TIP4P/2005). None of the MD force fields reproduces the amide I' VCD well but Amber ff14SB yields the least deviation from the experimental data. As reported in [Table S3](#) for alanine in GAG, Amber ff14SB predicts the pPII population of 0.55 in good agreement with the Gaussian model of 0.59. The corresponding pPII populations predicted by Amber ff99SBnmr1 (with SPC/E) and Amber ff03ws (with TIP4P/2005) of 0.44 and 0.41, respectively, are well below this value. OPLS-AA/L (with TIP4P) results in the lowest pPII population of 0.39, while the corresponding populations predicted by OPLS-AA/M (with TIP4P) and CHARMM36 both reach 0.51, in a relatively close agreement with the Gaussian model ([Table S3](#)). The close correspondence between OPLS-AA/M and Gaussian model with respect to pPII mesostate population might seem surprising considering a rather limited capability of OPLS-AA/M to reproduce the spectroscopic data. The reason for this contradiction can be inferred from the Ramachandran plot for OPLS-AA/M ([Figure 3](#)), which exhibits a pPII distribution that is shifted significantly to the left from the center of the pPII mesostate. While the probability density in the pPII basin is high enough to keep the population of the pPII mesostate high, this shift is large enough to cause the observed deviations from the experimental coupling constants and VCD profiles.

A significant difference between the MD-derived and Gaussian model Ramachandran distributions has to be addressed. For all six force fields, the peak of the β -strand population (i.e., a combined population of the $\alpha\beta$ and βt mesostates) lies within the $\alpha\beta$ mesostate, while in the Gaussian model, this population is centered in the βt region, that is, at less negative ϕ values. Intuitively, Ramachandran distributions

with a larger separation between pPII and β mesostates make physical sense because they are indicative of a significant barrier between the two states in the free energy landscape. We therefore asked to which extent the Gaussian model predictions would be affected if we shifted the β -strand population from the βt toward the $\alpha\beta$ mesostate. However, such a shift would require a simultaneous shift of the pPII population to maintain the ability of the Gaussian model to reproduce of two J coupling constants $^3J(\text{H}^{\text{N}}\text{H}^{\text{C}_\alpha})$ and $^3J(\text{H}^{\text{N}}\text{C}_\beta)$. This procedure would deteriorate the Gaussian model prediction of the other two 3J coupling constants and the VCD amide I' profile. We thus conclude that the Gaussian model parameters as reported in the previous work¹⁴ offer the optimal fit to the experimental data.

Conformational Ensembles of the Central Alanine in AAA. Force Field-Specific Ramachandran Plots for the Central Alanine in AAA. Ramachandran plots for the central alanine in AAA are obtained from MD simulations of cationic AAA in explicit water using the same six force fields as in the case of GAG ([Figure 3](#)). All MD-derived Ramachandran distributions for alanine in AAA in [Figure 5](#) demonstrate that the pPII population increases when terminal glycines are replaced by alanines, indicating that the nearest neighbor amino acids significantly affect conformational propensities as predicted by Toal et al.⁶⁸ This result is in a qualitative agreement with the predictions of the Gaussian model ([Figure 1a,b](#)), which also features an increased pPII population of alanine in AAA relative to GAG.

Mesostate populations are reported in [Table S5](#) for all six force fields and the predictions of the Gaussian model are included for comparison. Amber ff14SB again produces the highest pPII population (0.63) of all force fields under

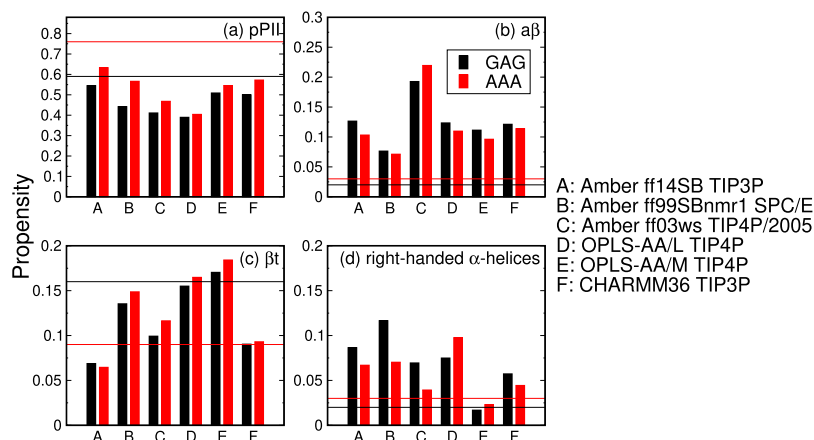


Figure 7. MD-derived mesostate populations for the central alanine in GAG and AAA. Black and red lines correspond to the mesostate populations for the central alanine in GAG and AAA, respectively, as predicted by the Gaussian model.

consideration, although the pPII basin in CHARMM36 (with TIP3P) is the most condensed in the (ϕ, ψ) space. The pPII population predicted by Amber ff14SB of 0.63 is lower than the pPII population of 0.76 predicted by the Gaussian model. Following Amber ff14SB are CHARMM36 and Amber ff99SBnmr1, which result in comparable pPII populations of ~ 0.57 , followed by OPLS-AA/M (0.55), Amber ff03ws (0.47), and OPLS-AA/L (~ 0.40). With respect to the $a\beta$ population, Amber ff03ws and Amber ff99SBnmr1 result in the highest and the lowest values, respectively. The remaining force fields produce comparable $a\beta$ populations, which are still higher than those predicted by the Gaussian model. OPLS-AA/M results in the highest βt population, followed by OPLS-AA/L and Amber ff99SBnmr1. Amber ff14SB is characterized by the lowest βt population. Whereas OPLS-AA/L produces the most populated α -helical mesostate, OPLS-AA/M results in the lowest α -helical population (<0.01) among the force fields under study. The discrepancies between the respective $a\beta$ and βt populations of the Gaussian model and the MD-derived distributions are again present. However, these discrepancies affect the J coupling constants and VCD profiles to a lesser extent than in the case of GAG because the overall β -strand sampling is decreased for the central alanine in AAA.

Amber ff14SB Outperforms the Other Force Fields in Capturing Conformational Ensembles of the Central Alanine in AAA. A comparison of MD-derived J coupling constants and amide I' VCD profiles to the respective experimental data, which is displayed in Figure 6, clearly indicates that Amber ff14SB (with TIP3P) again performs the best among the investigated force fields. Although the J coupling constants are slightly less well reproduced than they are in the case of GAG, ff14SB still results in the lowest χ_j^2 value. The J coupling constants calculated from simulations using different force fields and Gaussian model alongside experimental values and uncertainties used in χ_j^2 calculation are reported in Table S3. The corresponding VCD band profile is also quite close to the experimental profile. Interestingly, all force fields, in particular ff99SBnmr1 and ff03ws, much better reproduce the experimental VCD profile of the central alanine in AAA than the corresponding VCD profiles of alanine in GAG. The rather large χ_j^2 values (compared with GAG) are to a significant extent caused by the deviations between MD-derived and experimental $^3J(C, C')$ values. This coupling

constant is rather difficult to obtain, and the statistical error displayed in Figure 6f might not fully capture the uncertainties of the coupling constant and of the respective Karplus parameters. Moreover, all force fields overestimate the $^3J(H^N, H^{C\alpha})$ value when compared to the Gaussian model. Finally, the calculated $^3J(H^N, C_\beta)$ constants are all underestimated, even by the Gaussian model. This coupling constant is particularly tricky to handle. In order to obtain a value above 2 Hz, one has to consider some sampling in the upper right quadrant of the Ramachandran plot.^{12,13,16} Conformations in the upper right quadrant of the Ramachandran plot are absent from all MD-derived distributions as well as from the Gaussian model.

Comparison of Mesostate Populations of Alanine in GAG and AAA. A comparison of mesostate populations predicted by different force fields and the Gaussian model in Tables S3 and S5 and Figure 7 demonstrates that the propensity of alanine to adopt the pPII mesostate increases if the neighboring glycines (in GAG) are replaced by alanines (in AAA), which is consistent with the experimental data and indicative of the nearest neighbor interactions.⁶⁸

Thus, the pPII conformations dominate the conformational ensembles of alanine even more in AAA than in GAG. The GAG $\rightarrow$ AAA increase in the pPII population of $0.55 \rightarrow 0.63$ obtained for Amber ff14SB (with TIP3P) is quite close to the Gaussian model prediction of $0.59 \rightarrow 0.76$. The lowest pPII populations are predicted by OPLS-AA/L with the GAG $\rightarrow$ AAA increase of $0.39 \rightarrow 0.40$, followed by Amber ff03ws ($0.41 \rightarrow 0.47$), Amber ff99SBnmr1 ($0.44 \rightarrow 0.57$), OPLS-AA/M ($0.51 \rightarrow 0.55$), and CHARMM36 ($0.50 \rightarrow 0.57$). If the relative GAG $\rightarrow$ AAA increase in the pPII population is indicative of the strength of the nearest neighbor interactions, then these interactions are the strongest in Amber ff99SBnmr1 (1.30), followed by Amber ff14SB (1.15), Amber ff03ws (1.15), and CHARMM35 (1.14), whereas both OPLS force fields yield the lowest relative increases (1.08 and 1.03 for OPLS-AA-M and OPLS-AA-L, respectively).

It is noteworthy that $a\beta$ and βt populations are significantly less well accounted for by the six force fields under study, whereby $a\beta$ is significantly overpopulated for both GAG and AAA even in the best case (for Amber ff99SBnmr1 with SPC/E) relative to the corresponding population predicted by the Gaussian model. On the other hand, MD-derived βt populations are below the Gaussian model prediction of 0.16

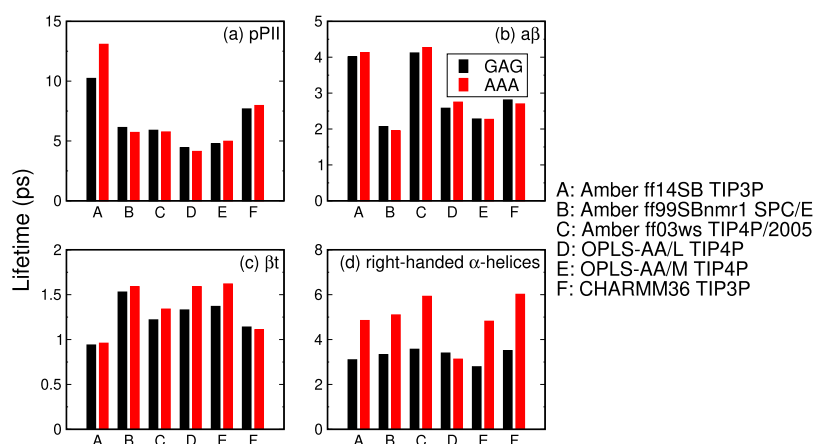


Figure 8. Lifetimes of the four mesostates for alanine in GAG and AAA derived from six different MD force fields.

Table 2. Mesostate Populations and Lifetimes of GAG and AAA in Amber ff99SB and Amber ff14SB

	pPII		$a\beta$		βt		right-handed α -helix	
	ff99SB	ff14SB	ff99SB	ff14SB	ff99SB	ff14SB	ff99SB	ff14SB
Populations								
GAG	0.31	0.55	0.20	0.13	0.13	0.07	0.07	0.09
AAA	0.45	0.63	0.21	0.10	0.16	0.07	0.04	0.07
Lifetimes [ps]								
GAG	4.53	10.23	3.08	4.01	1.32	0.94	1.96	3.10
AAA	5.37	13.08	3.43	4.13	1.44	0.96	3.18	4.85

for alanine in GAG for all force fields except for OPLS-AA/M, which predicts a higher value for this population than the Gaussian model. The reason for the increased βt population within OPLS-AA/M is the shift of the pPII basin to more negative ϕ values. However, in the Gaussian model βt decreases to 0.09 for alanine in AAA, whereas in all force fields other than Amber ff14SB (which predicts much lower βt populations than the Gaussian model), the βt population increases. Overall, Amber ff14SB outperforms the other five force fields with respect to its ability to reproduce the pPII mesostate populations for alanine in GAG and AAA. The next largest discrepancy between the Ramachandran distributions derived from Amber ff14SB (and other MD force fields under study) and those obtained in the Gaussian model revolve around the incorrect sampling of $a\beta$ and βt populations. Finally, although the right-handed α -helical mesostate has low occupancy in the Gaussian model, this mesostate remains oversampled in MD force fields other than OPLS-AA/M.

Lifetimes of Mesostates: A Glimpse into Their Stability. Here, we probe dynamic properties of the alanine residue within the six force fields under investigation by exploring the average time this residue remains in each mesostate. To this end, we calculate the mesostate lifetimes for the central alanine in both GAG and AAA, as described in [Methods](#). As shown in [Figure 8](#), the longest lifetimes are associated with pPII, which is also the most populated mesostate in alanine both in GAG and AAA within all six force fields. The lifetimes of $a\beta$ and right handed α helix are comparable but depend on the force field and are sensitive to nearest neighbors as GAG $\rightarrow$ AAA differences are notable. Lifetimes associated with βt are the shortest within all force fields.

The pPII lifetimes of different alanine's mesostates in GAG and AAA are significantly longer in Amber ff14SB than in the other five force fields. Because longer lifetimes of a particular

mesostate may be associated with a lower free energy and thus increased mesostate stability, the above observation indicates that Amber ff14SB stabilizes pPII better than the other force fields, most likely due to a decreased βt population, which produces a higher free energy barrier for pPII $\rightarrow$ β -strand transitions. Amber ff14SB and Amber ff03ws produce comparable $a\beta$ lifetimes, which are significantly longer than the corresponding lifetimes predicted by the other force fields. Of the six force fields, Amber ff14SB produces the lowest βt populations and the shortest βt lifetimes for alanine in GAG and AAA.

[Figure S10](#) shows that the pPII lifetime of the central alanine in AAA for Amber ff14SB, which was the longest recorded lifetime of all mesostates across all force fields under investigation, does not exceed 130 ps. Moreover, we calculated that more than 72% (for GAG) and 82% (for AAA) of pPII lifetimes are shorter than 20 ps. The corresponding average lifetimes of 10 ps (for GAG) and 13 ps (for AAA) are even shorter ([Table S6](#)). These results indicate that the total simulation time of 100 ns is 3 orders of magnitude longer than the longest mesostate lifetime, providing additional evidence that our MD simulations have converged.

Amber ff14SB³⁷ was developed from Amber ff99SB³² by recalculation of side-chain torsional potentials as well as by introducing changes to parameters associated with backbone dihedral potentials. Specifically, Maier et al. aimed to reduce sampling in the region between β and pPII (i.e., reduce the population of βt mesostate). To test to which extent these changes in the backbone dihedral potentials in Amber ff14SB affect mesostate populations and lifetimes, we performed MD simulations of GAG and AAA also within the canonical Amber ff99SB (combined with TIP3P). The corresponding mesostate populations and lifetimes for Amber ff14SB and Amber ff99SB are shown in [Table 2](#). In comparison to Amber ff99SB, Amber

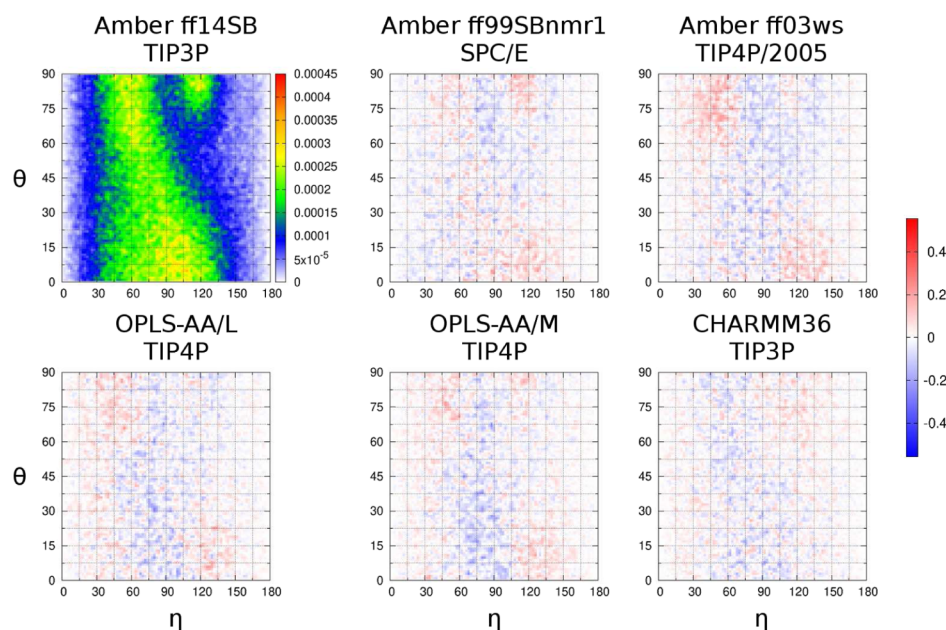


Figure 9. Difference in water orientation distributions between Amber ff14SB and other force fields for GAG. The top left water orientation plot corresponds to Amber ff14SB (with TIP3P). On the remaining plots, the water orientation distribution of Amber ff14SB is subtracted from the distribution in other force fields and the difference is normalized to the maximum value across all water orientation distributions (45×10^{-5}), such that the scale on the five difference plots corresponds to relative (fractional) changes.

ff14SB produces increased pPII populations for the central alanine in both GAG and AAA. These increased pPII populations are associated with significantly longer lifetimes, whereby the pPII lifetimes of alanine are increased by more than two-fold in both GAG and AAA. In contrast, Amber ff14SB-derived α/β populations of alanine in GAG and AAA are lower than the corresponding Amber ff99SB populations. Consistent with the reported reparameterization of the backbone dihedral potentials, Amber ff14SB predicts lower β t populations and shorter lifetimes than Amber ff99SB. These findings imply that the pPII mesostate in Amber ff14SB was stabilized by increasing its lifetime via increased free energy barrier for the pPII $\rightarrow$ β -strand transition.

GAG $\rightarrow$ AAA Increase in the pPII Population Correlates with the Changes in Water Orientation. We here asked if properties of water within the first hydration layer around the side chain of alanine are affected by the nearest neighboring amino acids. Because pPII conformations dominate the conformational dynamics of alanine, we derived water orientation plots as previously introduced by Meral et al.⁷ (see also [Methods](#)) for the ensemble of pPII conformations of alanine in GAG and AAA for all six force fields. The water orientation plot of alanine in GAG predicted by Amber ff14SB, shown in [Figure 9](#) (top leftmost plot), indicates that water around alanine's side chain adopts three predominant orientations: region A ($90^\circ < \eta < 100^\circ$, $\theta < 10^\circ$) wherein the water molecule plane is parallel to the SAS of the alanine's side chain; region B ($50^\circ < \eta < 70^\circ$, $\theta > 60^\circ$) wherein the water oxygen points toward and the two water hydrogens point away from the SAS of the alanine's side chain; and region C ($\eta \approx 120^\circ$, $80^\circ < \theta < 90^\circ$) wherein one of the two water hydrogens points toward and water oxygen points away from the SAS of the alanine's side chain in such a way that the normal to the water plane is nearly perpendicular to the SAS normal.

To examine the effect of the force field on the preferred water orientation next to the side chain of alanine in GAG, the water orientation distributions for the five remaining force fields are displayed in [Figure 9](#) as differences from the water orientation distribution predicted by Amber ff14SB (with TIP3P), whereby the scale on the five plots is normalized to show relative changes. These difference plots demonstrate that water orientation around the side chain is sensitive to the force field and water model as the average water orientation probability density changes up to $\pm 50\%$. Importantly, these variations are not random. For example, Amber ff99SBnmr1 (with SPC/E) predicts significantly more water orientations in region C than any other force field. Relative to Amber ff14SB (with TIP3P), in Amber ff99SBnmr1 (with SPC/E), water orientations in all three main regions A, B, and C are more populated at the expense of the more peripheral regions. One may argue that these differences stem from distinct water models. However, both Amber ff14SB and CHARMM36 use TIP3P yet CHARMM36 predicts more water orientations in region C at the expense of those in regions A and B (and in the transitional region between regions A and B). The two OPLS-AA force fields (both combined with TIP4P) also show distinct (nonrandom) patterns. Relative to Amber ff14SB, Amber ff03ws (with TIP4P/2005) exhibits a unique pattern predicting significantly more water orientations in region B as well as an increased population in region C, whereas the most populated region A is shifted to larger η values. These observations demonstrate that water orientation around alanine's side chain is strongly sensitive to the force field and water model.

Is the water orientation distribution around alanine's side chain sensitive to the nearest neighbor amino acids? To assess this sensitivity, we derive water orientation distributions around the side chain of the central alanine in AAA and compared them to the corresponding distributions for GAG. The results in [Figure 10](#) for Amber ff14SB (upper leftmost

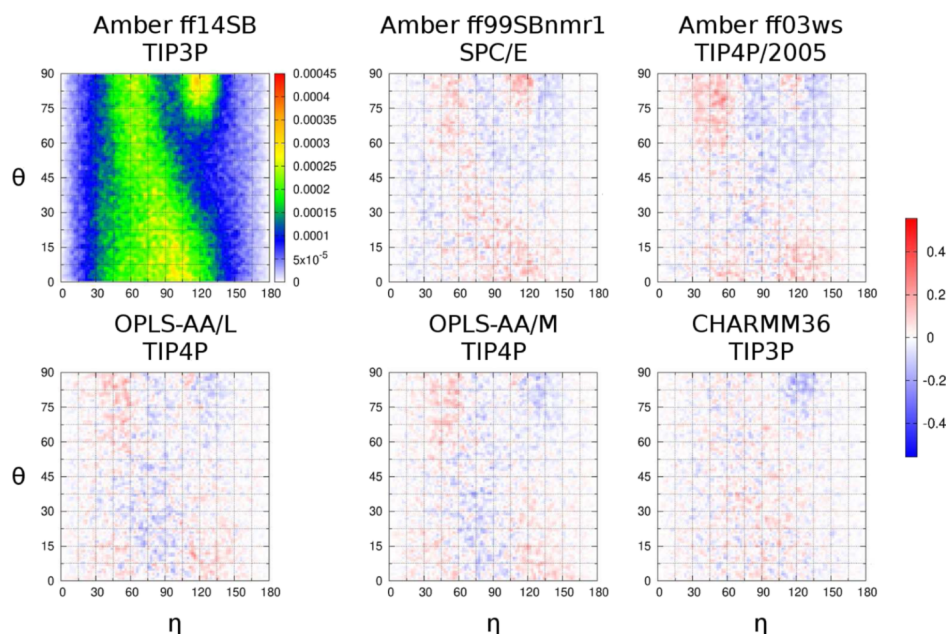


Figure 10. Difference in water orientation distributions between Amber ff14SB and other force fields for AAA. The top left water orientation plot corresponds to Amber ff14SB (with TIP3P). On the remaining plots, the water orientation distribution of Amber ff14SB is subtracted from the distribution in other force fields and the difference is normalized to the maximum value across all water orientation distributions (45×10^{-5}), such that the scale on the five difference plots corresponds to relative (fractional) changes.

plot) demonstrate that nearest neighbors have a significant effect on the orientation of hydration water adjacent to the side chain. This conclusion is valid for all six force fields as explicitly shown using the GAG $\rightarrow$ AAA difference distributions in Figure S12.

All six force fields predict increased populations of water orientations within region C whereby Amber ff14SB (with TIP3P) is associated with the largest and OPLS-AA/M (with TIP4P) with the lowest increase. Concomitant with the increase of water orientations in region C is a somewhat smaller decrease in water orientations in region A, again a feature shared across all six force fields. A comparison of the water orientation distribution around the side chain of the central alanine in AAA in Figure 10 to the one obtained for Amber ff14SB reveals that force field-specific differences follow similar force field-specific patterns as the ones for GAG (vide supra) except for CHARMM36. These differences are larger for the two Amber force fields. The water orientation distribution derived from CHARMM36 is more similar to the distribution obtained from Amber ff14SB than in the case of GAG (compare Figures 9 and 10, bottom right) with an exception of a decreased population of water orientations in region C.

Properties of Hydration Water around the Alanine's Side Chain. We here asked if the properties of hydration water (other than their orientation) adjacent to alanine's side chain are affected by the mesostate or by the nearest neighbors. We calculate the average number of water molecules and the average number of water–water hydrogen bonds formed by water in the hydration layer surrounding the side chain of the central alanine in GAG and AAA (see Methods). Figure S13 shows the average number of water molecules for each mesostate as predicted by the six force fields. On average, pII and right-handed α helix are associated with a larger number of water molecules around alanine's side chain than β t and α β , although the differences are in the order of a fraction of one

water molecule, that is, <10%. The most interesting is a slight but statistically significant decrease of $\sim 1/4$ in the average number of water molecules around alanine's side chain when glycines in GAG are replaced by alanines in AAA, which is observed for all six force fields. This result implies that the average of water molecules around the side chain of the central alanine residue is affected by its nearest neighbors. When combined with the above result on the changes in the average water orientation, these findings demonstrate that on average both the number and orientation of hydration water molecules adjacent to the alanine's side chain are sensitive to its nearest neighbors.

Figure S14 shows the average number of water–water hydrogen bonds that form in the hydration layer around the side chain of the central alanine. The four graphs in Figure S15 that correspond to different mesostates are self-similar, indicating that water–water hydrogen bonding is dominated by the water model such that force fields combined with the same water model (Amber ff14SB and CHARMM36 are both combined with TIP3P whereas both OPLS-AA force fields use TIP4P) predict similar average numbers of water–water hydrogen bonds. The average number of water–water hydrogen bonds in the hydration layer around the side chain of alanine is the highest for SPC/E and TIP4P/2005, somewhat lower for TIP4P, and the lowest for TIP3P, showcasing the structural properties of water as predicted by different water models.^{69,70} Within each of the six force fields, the average number of water–water hydrogen bonds is larger in pII than in α β for both GAG and AAA. No such conclusion can be made when pII is compared to β t or right-handed α helix. The average number of water–water hydrogen bonds around the central alanine tends to decrease when the neighboring glycines (GAG) are replaced by alanines (AAA) for all six force fields; however, this decrease is not statistically significant for all force fields.

Properties of Hydration Water around the Alanine's Backbone. To complement Figure S13, which shows the average number of water molecules in the hydration layer around the side chain of alanine in GAG and AAA, Figure S15 shows the same quantity calculated for the hydration layer around the backbone of alanine in GAG and AAA as predicted by the six force fields. The results in Figure S15 show that pPII and right-handed α helix are associated with on average higher numbers of water molecules in the hydration layer than $\alpha\beta$ and $\beta\beta$, similar to the case of the side chain hydration layer (Figure S13). When glycines in GAG are replaced by alanines, Figure S13 reveals a smaller average number of water molecules around the side chain of the central alanine. One might anticipate that this small ($\sim 1/4$ of a water molecule) but statistically significant deficit of water molecules would be compensated for in the hydration layer around the alanine's backbone. This prediction would be incorrect as clearly displayed in Figure S15, which demonstrates an even larger decrease of ~ 1 in the average number of water molecules in the hydration layer around the alanine's backbone in AAA relative to GAG.

Is the number of water–alanine hydrogen bonds affected by the mesostate or nearest neighbors? To answer this question, we calculated the average number of hydrogen bonds between water and the central alanine residue in GAG and AAA for all six force fields. Because the side chain of alanine does not form hydrogen bonds with water (data not shown), the number of water–peptide hydrogen bonds displayed in Figure S16 reflects the number of hydrogen bonds between water and the functional (NH and CO) groups of alanine. We note that the average number of water–alanine hydrogen bonds is significantly less affected by the water model than the average number of water–water hydrogen bonds in the hydration layer around alanine's side chain (Figure S14). For all six force fields, the highest number of water–peptide hydrogen bonds is formed when alanine adopts pPII. With respect to nearest neighbor effects, Amber ff03ws (with TIP4P/2005) is the only force field that predicts a decreased average number of water–alanine hydrogen bonds when glycines (in GAG) are replaced with alanines (in AAA), regardless of the mesostate. For the other five force fields (and their respective water models), the water–alanine hydrogen bonding is not strongly affected by nearest neighbors as both central alanines in GAG and AAA form on average comparable numbers of water–alanine hydrogen bonds. Although the hydration layer around the alanine's backbone in AAA contains on average fewer water molecules than in GAG (as predicted by all force fields in our study), the water–alanine hydrogen bonding is not affected except for Amber ff03ws.

The above results indicate that the hydration layer around the entire alanine residue contains on average a fewer number of water molecules when the neighboring glycines are replaced by alanines and the effect is significantly stronger in the hydration layer surrounding its backbone atoms. We then asked whether the total number of water–water hydrogen bonds around the entire alanine is also affected by the neighboring amino acids. Figure S16 confirms that this is indeed the case as the average number of water–water hydrogen bonds around the entire alanine residue significantly decreases when glycines (in GAG) are replaced by alanines (in AAA) within all force fields. This decrease is particularly large for Amber ff03ws (with TIP4P/2005), whereby the decrease in the average number of water–water hydrogen bonds is ~ 4 .

Notably, just like in the case of water–water hydrogen bonding in the hydration layer adjacent to the side chain (Figure S14), the average number of water–water hydrogen bonds around the entire alanine is mostly dominated by the structural properties of the water model with SPC/E showing the largest numbers, followed by TIP4P/2005, TIP4P, and TIP3P with the lowest number of water–water hydrogen bonds. The observation that fewer water molecules in the hydration layer around alanine in AAA relative to GAG are concomitant with fewer water–water hydrogen bonds in this layer may not be surprising. However, the observation that the hydration layer around the alanine's backbone rather than the side chain is the main contributor to these differences is unexpected as it showcases strong water-mediated interactions between alanine's side chain and backbone dynamics as well as between alanine and its nearest neighbors.

CONCLUSIONS

MD simulations provide atomistic insights into the dynamics of protein folding and protein–protein interactions. The predictive power of MD, however, depends on the accuracy of the underlying force field. Some deficiencies of commonly used biomolecular MD force fields are well-documented.^{4,5,71–73} One of the fundamental challenges of any MD force field is to predict intrinsic conformational propensities and dynamics of guest amino acids x in GxG peptides in water in a way that is consistent with the published experimental data.⁷ Of all naturally occurring amino acids, alanine has drawn a lot of attention in the literature because of its high propensity to adopt pPII (see reviews^{74,75} and references therein). Alanine-based short peptides are important as they are often used as reference systems for MD force field parameterization.

In this work, we examine several MD force fields with respect to their ability to capture experimentally determined J coupling constants and amide I' profiles for the central alanine in GAG and AAA. We explore Amber ff14SB (with TIP3P), Amber ff99SBnmr1 (with SPC/E), Amber ff03ws (with TIP4P/2005), CHARMM36 (with TIP3P), and both OPLS-AA force fields, OPLS-AA/L and OPLS-AA/M. For each force field, MD-derived Ramachandran distributions are used to calculate five (in the case of GAG) or six (in the case of AAA) J coupling constants and VCD amide I' profile to facilitate a direct comparison to spectroscopic data. As a benchmark, we also use the predictions of the Gaussian model, which employs a sum of symmetric 2D Gaussian subdistributions to model the Ramachandran distribution, whereby the parameters of the Gaussian subdistributions are optimized to best fit the experimental data. The overall performance of each force field and Gauss's model is then assessed via χ_f^2 and χ_{VCD}^2 values. Although we define four mesostates as rectangular regions in the Ramachandran space as visual guides, our assessment of force fields (alongside the respective water models) is independent of the definition of mesostates because both χ^2 functions are calculated using the entire Ramachandran distributions and not only the regions enclosed by the four mesostates.

For both OPLS-AA force fields, we first examine the effect of a water model on the Ramachandran distributions of alanine in GAG, using SPC/E, TIP3P, TIP4P, and TIP4P/2005, identifying TIP4P as the optimal choice. Examining the performance of the six force fields with respect to their ability to capture experimental data for the central alanine in GAG

and AAA shows that Amber ff14SB outperforms the other five force fields, giving rise to pPII populations for alanine in GAG and AAA that are in good agreement with those predicted by the Gaussian model. Two other force fields, Amber ff99SBnmr1 (with SPC/E) and CHARMM36 (with TIP3P), produce fairly good results. Amber ff03ws (with TIP4P/2005) and OPLS-AA/L (with TIP4P) show rather strong deviations from the experimental data.

Although the new version of the OPLS-AA force field, OPLS-AA/M (with TIP4P) is strongly improved over OPLS-AA/L (with TIP4P) in predicting pPII populations in GAG and AAA; these two force fields performed poorly in comparison to the others in this study, in particular with respect to predictions of two J coupling constants, $^3J(\text{H}^{\text{N}}, \text{H}^{\text{C}_\alpha})$ and $^3J(\text{H}^{\text{N}}, \text{C}')$, which may be in part associated with a shift of pPII populations to more negative ϕ values in the corresponding Ramachandran distributions of alanine in both GAG and AAA. To make this assessment more quantitative, we shifted the entire OPLS-AA/M (with TIP4P) dihedral angle distributions for GAG and AAA by 1 and 10°, respectively, in the positive ϕ angle direction, to match the centers of the pPII populations in the respective Gaussian models. We then calculated the J coupling constants and amide I' profiles for thus shifted OPLS-AA/M distributions and calculated the respective χ^2 and χ_{VCD}^2 values. For GAG, the 1° shift resulted in a minor decrease in χ^2 and χ_{VCD}^2 from 29.50 to 27.85 and from 11.03 to 10.85, respectively. For AAA, the 10°-shift resulted in a decrease in χ^2 and χ_{VCD}^2 from 59.76 to 31.44 and from 5.42 to 1.77, respectively. While not sufficient to outperform Amber ff14SB, thus shifted OPLS-AA/M would become more comparable in performance to CHARMM36 for the central alanine in AAA.

The success of Amber ff14SB in the description of alanine conformations represents a great improvement over the canonical Amber ff99SB force field, which strongly underestimates pPII populations in GAG and AAA alike. This improvement most likely stems from the empirical adjustments of dihedral potentials for ϕ and ψ , which reduced the population of the βt mesostate, thereby stabilizing pPII by increasing the free energy barrier for pPII/ β -strand transitions, which is reflected also in a significant increase in pPII lifetimes for both GAG and AAA. Paradoxically, this reduction in the βt population makes Amber ff14SB-derived Ramachandran distributions for GAG and AAA more distinct from the predictions of the Gaussian model. In fact, all MD force fields under study exhibit similar deviations from the Gaussian model-derived Ramachandran distributions of the central alanine in both GAG and AAA: (i) the pPII basin predicted by the Gaussian model is significantly broader than all MD-derived pPII basins; (ii) all MD force fields predict overpopulated $\alpha\beta$ and underpopulated βt mesostates; and (iii) there are two populations in the Ramachandran distributions of alanine in GAG (see Figure 1a, γ turn population) and AAA (see Figure 1b, inverse γ turn population) derived from the Gaussian model, which are not accounted for by MD force fields under investigation. Our findings thus suggest three potential ways to improve the ability of MD force fields to capture the experimental data.

Although Amber ff14SB (with TIP3P) outperforms the other force fields in predicting alanine conformational dynamics in water, CHARMM36 (with TIP3P) performs quite well and does not lag behind Amber ff14SB very much. Importantly, there is no guarantee that Amber ff14SB will

perform equally well when predicting conformational dynamics of other amino acid residues in water. Nonetheless, studies of alanine and alanine-based peptides help elucidate the peculiarly high propensity of this amino acid residue to adopt the pPII mesostate. Meral et al. proposed that pPII conformations of amino acid residues in water are stabilized by the clathrate-like water structure around them.⁷ We here further explore properties of hydration water around alanine in GAG and AAA for each of the four mesostates. Of the four mesostates, pPII emerges as the state, in which water forms the most hydrogen bonds with the functional groups of the alanine's backbone, which may anchor the water clathrate structure to alanine. Our analysis of orientational degrees of hydration water shows that in all six MD force fields under study, the average orientation of water molecules in the hydration layer around the side chain of alanine in the pPII state is significantly affected by alanine's nearest neighbors. Namely, replacing glycines (in GAG) by alanines (in AAA) increases the relative population of water orientations, in which one of the two water hydrogens is turned toward the SAS of the side chain. This water orientation is typical for amino acid side chains that engage in hydrogen bonding with water, which is surprising because the alanine's side chain is not involved in the hydrogen bond formation. Concomitant with this GAG $\rightarrow$ AAA water reorientation in the hydration layer adjacent to alanine's side chain is a loss of on average >1 water molecule from the hydration layer around alanine. This loss of hydration water around alanine in AAA relative to GAG is accompanied by a loss of water–water hydrogen bonding in the hydration layer surrounding the entire alanine, which is particularly large in the hydration layer adjacent to the alanine's backbone. These results combined provide a direct atomistic evidence of hydration water-mediated interactions between the alanine's side chain and backbone groups, a phenomenon that is relevant to understanding conditional solvation.⁷⁶ Moreover, our findings demonstrate that several hydration water properties are strongly sensitive to neighboring amino acid residues, elucidating the molecular basis of nearest neighbor effects in short unfolded peptides in water.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.9b00588>.

Mesostate populations from different force fields and Gaussian model for GAG and AAA; p_0 , τ , and R^2 values from fitting of mesostate lifetime distributions of pPII, βt , $\alpha\beta$, and right-handed α -helix of GAG and AAA; four J coupling constants: $^3J(\text{H}^{\text{N}}, \text{H}^{\text{C}_\alpha})$, $^3J(\text{H}^{\text{N}}, \text{C}')$, $^3J(\text{H}^{\text{N}}, \text{C}_\beta)$, and $^3J(\text{H}^{\text{C}_\alpha}, \text{C}')$ as functions of ϕ ; sensitivity analysis of pPII within the Gaussian model for GAG; comparison among water models combined with OPLS-AA/L and OPLS-AA/M on GAG; time evolution of Ramachandran plots for alanine in GAG; time evolution of J coupling constants of alanine in GAG; time evolution of mesostate populations of alanine in GAG; amide I' profiles of alanine in GAG; mesostate lifetime distributions of GAG in Amber ff14SB; AAA to GAG difference in water orientation in the hydration layer of the central alanine side chain; number of water molecules and number of water–water hydrogen bonds in the hydration layer of the central alanine side

chain; number of water molecules in the hydration layer of the central alanine backbone; hydrogen bonds between water and the central alanine; and water–water hydrogen bonds in the hydration layer of the central alanine (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: bu25@drexel.edu. Fax: (215) 895-5934.

ORCID

Reinhard Schweitzer-Stenner: 0000-0001-5616-0722

Brigita Urbanc: 0000-0001-9159-7698

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation grant MCB-181765.

REFERENCES

- (1) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331.
- (2) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26–59.
- (3) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
- (4) 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.
- (5) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95*, L07–L09.
- (6) Toal, S.; Meral, D.; Verbaro, D.; Urbanc, B.; Schweitzer-Stenner, R. The pH-independence of trialanine and the effects of termini blocking in short peptides: A combined vibrational, NMR, UVCD, and molecular dynamics study. *J. Phys. Chem. B* **2013**, *117*, 3689–3706.
- (7) Meral, D.; Toal, S.; Schweitzer-Stenner, R.; Urbanc, B. Water-centered interpretation of intrinsic pPII propensities of amino acid residues: In vitro-driven molecular dynamics study. *J. Phys. Chem. B* **2015**, *119*, 13237–13251.
- (8) Grdadolnik, J.; Mohacek-Grosov, V.; Baldwin, R. L.; Avbelj, F. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1794–1798.
- (9) He, L.; Navarro, A. E.; Shi, Z.; Kallenbach, N. R. End Effects Influence Short Model Peptide Conformation. *J. Am. Chem. Soc.* **2012**, *134*, 1571–1576.
- (10) Schweitzer-Stenner, R. Distribution of Conformations Sampled by the Central Amino Acid Residue in Tripeptides Inferred From Amide I' Band Profiles and NMR Scalar Coupling Constants. *J. Phys. Chem. B* **2009**, *113*, 2922–2932.
- (11) Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. Intrinsic Propensities of Amino Acid Residues in GxG Peptides Inferred from Amide I' Band Profiles and NMR Scalar Coupling Constants. *J. Am. Chem. Soc.* **2010**, *132*, 540–551.
- (12) Hagarman, A.; Mathieu, D.; Toal, S.; Measey, T. J.; Schwalbe, H.; Schweitzer-Stenner, R. Amino Acids with Hydrogen-Bonding Side Chains have an Intrinsic Tendency to Sample Various Turn Conformations in Aqueous Solution. *Chemistry* **2011**, *17*, 6789–6797.
- (13) Rybka, K.; Toal, S. E.; Verbaro, D. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. Disorder and order in unfolded and disordered peptides and proteins: A view derived from tripeptide conformational analysis. II. Tripeptides with short side chains populating α and β -type like turn conformations. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 968–983.
- (14) Schweitzer-Stenner, R.; Hagarman, A.; Toal, S.; Mathieu, D.; Schwalbe, H. Disorder and order in unfolded and disordered peptides and proteins. A view derived from tripeptide conformational analysis. I. Tripeptides with long and predominantly hydrophobic side chains. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 955–967.
- (15) Toal, S. E.; Verbaro, D. J.; Schweitzer-Stenner, R. Role of Enthalpy-Entropy Compensation Interactions in Determining the Conformational Propensities of Amino Acid Residues in Unfolded Peptides. *J. Phys. Chem. B* **2014**, *118*, 1309–1318.
- (16) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
- (17) Schweitzer-Stenner, R. Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure - A critical review. *Vib. Spectrosc.* **2006**, *42*, 98–117.
- (18) DiGiuseppi, D.; Schweitzer-Stenner, R. Probing conformational propensities of histidine in different protonation states of the unblocked glycyl-histidyl-glycine peptide by vibrational and NMR spectroscopy. *J. Raman Spectrosc.* **2016**, *47*, 1063–1072.
- (19) 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.
- (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) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
- (22) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free-energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. *J. Phys. Chem.* **1994**, *98*, 13077–13082.
- (23) 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.
- (24) 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*, 6474–6487.
- (25) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509.
- (26) 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.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (27) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (28) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (29) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of

proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(30) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(32) 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: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(33) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(34) Li, D.-W.; Brüschweiler, R. NMR-Based Protein Potentials. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 6778–6780.

(35) 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: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.

(36) Zhou, C.-Y.; Jiang, F.; Wu, Y.-D. Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER ff99SB. *J. Phys. Chem. B* **2014**, *119*, 1035–1047.

(37) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(38) Pang, Y.-P. FF12MC: A revised AMBER forcefield and new protein simulation protocol. *Proteins: Struct., Funct., Bioinf.* **2016**, *84*, 1490–1516.

(39) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martinez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023–4039.

(40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. , <http://www.rcsb.org/pdb>

(41) 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*, 14916–14923.

(42) 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*, 5113–5124.

(43) Pappu, R. V.; Srinivasan, R.; Rose, G. D. The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565–12570.

(44) Tran, H. T.; Wang, X.; Pappu, R. V. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* **2005**, *44*, 11369–11380.

(45) Li, S.; Andrews, C. T.; Frembgen-Kesner, T.; Miller, M. S.; Siemonsma, S. L.; Collingsworth, T. D.; Rockafellow, I. T.; Ngo, N. A.; Campbell, B. A.; Brown, R. F.; Guo, C.; Schrodt, M.; Liu, Y.-T.; Elcock, A. H. Molecular Dynamics Simulations of 441 Two-Residue Peptides in Aqueous Solution: Conformational Preferences and Neighboring Residue Effects with the Amber ff99SB-ildn-NMR Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 1315–1329.

(46) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190–9195.

(47) Ding, L.; Chen, K.; Santini, P. A.; Shi, Z.; Kallenbach, N. R. The pentapeptide GGAGG has PII conformation. *J. Am. Chem. Soc.* **2003**, *125*, 8092–8093.

(48) Chen, K.; Liu, Z.; Kallenbach, N. R. The polyproline II conformation in short alanine peptides is noncooperative. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15352–15357.

(49) 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.

(50) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.

(51) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(52) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(53) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(54) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. Solving software challenges for exascale. *2nd International Conference on Exascale Applications and Software (EASC), Stockholm, SWEDEN, APR 02-03, 2014, 2015*; pp 3–27.

(55) 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.

(56) Páll, S.; Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **2013**, *184*, 2641–2650.

(57) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(58) 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.

(59) Hu, J.-S.; Bax, A. Determination of ϕ and χ_1 angles in proteins from ^{13}C – ^{13}C three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J. Am. Chem. Soc.* **1997**, *119*, 6360–6368.

(60) Wang, A. C.; Bax, A. Determination of the backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical Karplus equations. *J. Am. Chem. Soc.* **1996**, *118*, 2483–2494.

(61) Case, D. A.; Scheurer, C.; Brüschweiler, R. Static and dynamic effects on vicinal scalar J couplings in proteins and peptides: A MD/DFT analysis. *J. Am. Chem. Soc.* **2000**, *122*, 10390–10397.

(62) Wirmer, J.; Schwalbe, H. Angular dependence of $^1J(N_\nu C_{\alpha i})$ and $^2J(N_\nu C_{\alpha(i-1)})$ coupling constants measured in J-modulated HSQC. *J. Biomol. NMR* **2002**, *23*, 47–55.

(63) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(64) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(65) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95.

(66) Brant, D. A.; Flory, P. J. The Configuration of Random Polypeptide Chains. II. Theory. *J. Am. Chem. Soc.* **1965**, *87*, 2791–2800.

- (67) Feig, M. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *J. Chem. Theory Comput.* **2008**, *4*, 1555–1564.
- (68) Toal, S. E.; Kubatova, N.; Richter, C.; Linhard, V.; Schwalbe, H.; Schweitzer-Stenner, R. Randomizing the unfolded state of peptides (and proteins) by nearest neighbor interactions between unlike residues. *Eur. J. Chem.* **2015**, *21*, 5173–5192.
- (69) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragonés, J. L. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss.* **2009**, *141*, 251–276.
- (70) Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663–19688.
- (71) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47–L49.
- (72) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PLoS One* **2012**, *7*, No. e32131.
- (73) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409–1414.
- (74) Schweitzer-Stenner, R. Conformational propensities and residual structures in unfolded peptides and proteins. *Mol. BioSyst.* **2012**, *8*, 122–133.
- (75) Toal, S.; Schweitzer-Stenner, R. Local order in the unfolded state: Conformational biases and nearest neighbor interactions. *Biomolecules* **2014**, *4*, 725–773.
- (76) Ben-Naim, A. *Molecular Theory of Water and Aqueous Solutions—Part I: Understanding Water*; World Scientific: Singapore, 2009.