Article

Development and Validation of a DFT-Based Force Field for a Hydrated Homoalanine Polypeptide

Ying Yuan, Zhonghua Ma, and Feng Wang*

Cite This: *J. Phys. Chem. B* 2021, 125, 1568−1581

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations | 🆂🅸 Supporting Information

ABSTRACT: A new force field has been created for simulating hydrated alanine polypeptides using the adaptive force matching (AFM) method. Only density functional theory calculations using the Perdew−Burke−Ernzerhof exchange−correlation functional and the D3 dispersion correction were used to fit the force field. The new force field, AFM2020, predicts NMR scalar coupling constants for hydrated homopolymeric alanine in better agreements with experimental data than several other models including those fitted directly to such data. For Ala_7, the new force field shows about 15% helical conformations, 20% conformation in the β basin, and 65% polyproline II. The predicted helical population of short hydrated alanine is higher than previous estimates based on the same experimental data. Gas-phase simulations indicate that the force field developed by AFM solution-phase data is likely to produce a reasonable conformation distribution when hydration water is no longer present, such as the interior of a protein.
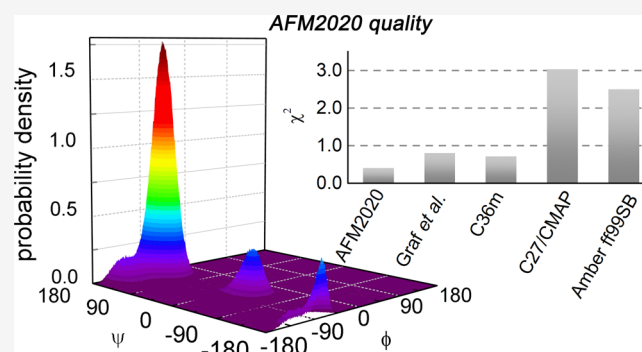
1. INTRODUCTION

All-atom simulations are an important technique in molecular biology. While experimental techniques, such as X-ray crystallography, provide a static picture, simulations provide mechanistic and dynamic insights on important biological processes. Interpretation of experimental data frequently relies on modeling.[1−5] Computational biology has contributed to our understanding of many fundamental biological processes and is now considered essential in many areas of biomedical research, such as computer-aided drug design.[6]

A trustworthy molecular simulation requires an accurate potential that describes atomistic interactions. Although *ab initio* equations can be solved to provide such an interaction potential, the vast majority of computational biology relies on a simplified set of molecular mechanics equations with associated parameters to describe the atomistic scale interactions. Such a set of equations along with the parameters is widely referred to as a force field.

Development of biological force fields, in particular, protein force fields, has a long history.[7−10] A protein force field has a large number of parameters. The majority of the force fields use simple Lennard-Jones terms to describe repulsion and dispersion, although other power-law-based terms are also used.[11,12] We will refer to the combination of repulsion and dispersion as short-range nonbonded interactions. Generally, a protein force field derives the nonbonded parameters and partial charges from gas-phase electronic structure calculations. For example, CHARMM fits the nonbonded interactions based on gas-phase energies of a model compound and water.[13] The

partial charges can be obtained by fitting the electrostatic potential of the model compound in vacuum. Such a fit typically relies on Hartree−Fock calculations with additional restraints, such as the RESP algorithm.[14] The charges so obtained are frequently scaled to compensate for enhanced polarizations in the condensed phase. More recent approaches include extrapolation of partial charges fitted in vacuum and in implicit solvents.[15,16] Intramolecular bond, angle, and torsional terms generally rely on experimental data on model compounds or electronic structure-based energy surfaces.[13,15,17−19] For example, the torsional terms are frequently fitted to the local MP2 potential energy surfaces in the gas phase. In many cases, electronic structure-based torsional terms are found to be less than satisfactory; it is sometimes necessary to adjust the torsional terms empirically to improve agreement on selected proteins.[20]

Many protein potentials have several sets of parameters in use.[16,20−28] Some protein force fields have multiple lineages.[16,23,24] When the parameter sets are being revised, frequently only the torsional terms are revised with the partial
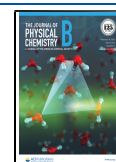
charges and nonbonded short-range interactions inherited from earlier models in the same lineage.[20,22,23,25,26,28,29]

In this publication, we present the development of an alanine force field based on and only on condensed phase force-matching.[30−32] Partial charges, short-range repulsion, and torsional surfaces are all determined directly in the condensed phase without assuming transferability between gas phase and condensed phase. It is well known that many-body effects play an important role in a condensed phase.[33] For pair-wise additive potentials, we believe that direct condensed phase fitting is especially meritorious since it allows additive expressions to better capture many-body effects implicitly.

We will focus on developing a pair-wise additive potential, which allows inexpensive simulations at the microsecond scale to be performed. We will leverage the adaptive force matching (AFM) method,[34−36] which provides a mechanism to fit condensed phase reference forces on the most thermodynamically important conformations.

We aim to create a whole new force field with new definitions of atom types. All related parameters, from partial charges and nonbonded short-range interactions to intramolecular bond, angle, and dihedral terms, are fitted. In the future, if additional training conformations are deemed important, the reference forces released with the publication will allow all parameters to be refit self-consistently with the create-your-own-force-field (CRYOFF) code.[37] Such a refit of all parameters is better than only patching part of the energy expressions, such as torsional terms, to fix problems. Rather than relying on Lennard-Jones terms, we will use the Buckingham potential[38] since its exponential repulsion is expected to be more physical.

Leveraging the ability of AFM to reliably obtain a large number of parameters, we will not use combination rules. Nonbonded short-range parameters between any unique pairs of atom types will be fitted. The α-helix is expected to be stabilized by the intrapeptide, NH···OC hydrogen bond. When relying on combination rules, parameters have to reach a compromise between modeling hydration effects dictated by amide and carbonyl−water interactions and modeling the NH···OC hydrogen bond. In fact, many models fit the amide−water and carbonyl−water interactions to obtain the NH and OC parameters and rely on combination rules for modeling the intrapeptide hydrogen bond.[13,17,18] By not using combination rules, short-range nonbonded terms can be optimized for modeling the peptide−peptide and peptide−water interactions independently.

Alanine was chosen in this work as there is a rich body of both theoretical and experimental work on homologous alanine.[39−46] It is considered one of the best α-helix formers,[42−45] and at the same time, hydrated short alanine is an example of an intrinsically disordered protein.[47] In this work, we will focus on hydrated homologous alanine peptides of various lengths. While the development of the force field relies completely on electronic structure based force matching (FM),[34] the validation of the fitted model will be performed by comparing with scalar coupling constants from two-dimensional NMR spectra.[40] The scalar coupling data provide insights into conformation distributions of the peptide secondary structures through the Karplus equation.[48] While such data sets have been used to reparameterize torsional terms in the existing force fields in various studies,[25,49,50] we will use such couplings only as a validation. The performance of our electronic structure-based force field for reproducing the J-coupling will be compared to

other known models including models where the J-coupling has been fitted directly.

It is found that our AFM-based alanine model outperforms even a direct thermodynamic fit to the experimental J-coupling constants, strongly endorsing the quality of our force field for modeling hydrated alanine peptide. Although our force field was fitted to reference forces of a hydrated peptide, simulations were also performed in the absence of hydration water. A reasonable conformation distribution was obtained, suggesting that our force field is applicable to a broader range of systems with alanine. We note that our model is just a small step toward a full protein force field based on AFM. The performance of our model for modeling a real protein has not been fully evaluated.

The paper is composed of the following sections. After the introduction, a discussion of the AFM-based alanine force field development will be presented in Section 2. Section 3 reports computational details of our validation. A quality assessment of our potential is presented in Section 4. A comparison between our model and selected models from the AMBER and CHARMM family is presented in Section 5. A summary and discussion are presented in Section 6.

2. ADAPTIVE FORCE MATCHING

The AFM method was originally developed in the Wang group.[34−36,51] Although the method has been shown to give quantitatively accurate potentials for many simple systems, such as water,[51,52] hydrated ions,[53] and CO_2,[54] its application to more complex systems has not been explored. Peptide and protein are challenging test cases since an error of 1 kcal/mol is expected to change the relative population of different conformations by a factor of 10 based on the Boltzmann law.

In this work, we developed a force field for poly-alanine peptide. A typical AFM procedure iterates through three steps, the molecular dynamics (MD) step, the quantum mechanics (QM)/molecular mechanics (MM) step, and the FM step. A detailed description of the AFM method has been provided previously.[51] We will only focus on describing procedures specific to the fitting of the alanine force field.

The fitting will be performed using hydrated Ala_7 as the model system. The choice reflects a compromise between the computational cost of the required QM/MM calculations and the proper sampling of peptide hydrogen bonds that are responsible for the stability of α-helix. Most biological textbooks attribute the stability of the α-helix to the intrapeptide hydrogen bond between residues i and $i + 4$, although some force fields have the α-helix to be a minimum in the torsional energy surface, thus capable of forming α-helices even without such hydrogen bonds. We anticipate that it is important to ensure sampling of i and $i + 4$ hydrogen bonds in our training set. To form a $i \rightarrow i + 4$ hydrogen bond, the shortest peptide would be Ala_5. However, we decided to use a zwitterionic peptide, leading to the two terminal residues to have different parameters. It was thus decided that Ala_7 should be sufficiently long but not so long as to make the QM/MM computations impractical.

Since we intended to use the NMR scalar coupling data to validate the final force field, a zwitterionic system would be most similar to the cationic system used for the NMR study without the need to parameterize a counterion.

In this work, only the alanine−water and alanine−alanine parameters will be optimized using AFM. We will use the BLYPSP-4F model developed previously with AFM for modeling water.[51] This model gives many properties of water in excellent agreement with experiments, such as radial

distribution functions (RDFs), diffusion constant, dielectric constant, and heat of vaporization.[55] Previous studies also showed that the hydration free energies of salts[53] and small solute molecules[56] were in good agreement with experiments when fitted with this water model using AFM.

The Amber ff99SB force field[21] will be used as the initial guess for the first generation of AFM. Water–water interactions are modeled with BLYPSP-4F. Since BLYPSP-4F does not use Lennard-Jones for nonbonded short-range interactions, to work with ff99SB as the initial guess force field, the alanine–water cross-terms were derived by assuming that BLYPSP-4F water has the same σ and ε parameters as TIP3P.[57] We note that combination rules are only used to construct the initial guess and are not needed after the first generation.

2.1. MD Step. The AFM method obtains its training set conformations with the MD sampling step. In this work, the MD sampling was performed with one zwitterionic alanine–heptamer and 3021 water molecules. The box is large enough to allow for approximately 1.0 nm of water to surround the alanine–heptamer on all sides, even in the most stretched conformation.

Although alanine is known to be one of the best α-helix formers, shorter chains may not have much α-helix conformation. Also, β-sheets are stabilized by hydrogen bonding between peptide strands. A short peptide will be difficult to fold back to form "interchain" β-sheets and would have less β-sheet conformations. Since we hope our force field will also perform well for much longer alanine peptides and be used as part of a protein force field, we decided to generate restrained training sets to ensure sampling of such important regions of the conformation space.

To accomplish this and ensure balanced sampling, we include five groups of conformations by performing separate MD simulations. All torsional restraints were enforced using a cosine dihedral formula (see eq 3 later) with $n = 1$ and a force constant of 50 kJ/mol unless specified otherwise.

(a) α-Helix group: the ϕ and ψ angles were restrained to -60 and $-45°$, respectively, to sample the α-helix basin. In addition, we added harmonic restraint of $i, i + 4$ hydrogen bonds to 1.9 Å with a force constant of 1000 kJ/(mol·Å^2), and the N–H–O angle of the hydrogen bond to 180° with a force constant of 500 kJ/(mol·rad^2).

(b) β-Strand group: the ϕ and ψ angles were restrained to -135 and $135°$, respectively.

(c) Polyproline II (PPII) group: the ϕ and ψ angles were restrained to -75 and $150°$, respectively.

(d) Survey group: with only the α-helix, β-strand, and PPII restraints, the torsional space will be under-represented, which could lead to stability issues for other regions of the torsional surface. To cover a wider range of torsional angles in the training set, the survey group was created. This group contains four subsets each with a separate restraint. The restraints were set at ($\psi = -120°$), ($\phi = -150°, \psi = -60°$), ($\phi = -150°, \psi = 60°$), and ($\phi = -90°, \psi = 60°$). These restraints were enforced with a smaller force constant of 20 kJ/mol to allow for a wider distribution around these points.

(e) Unrestrained group: we also performed MD for Ala$_7$ without any restraint to provide an unbiased training set.

All sampling was performed in the *NPT* ensemble at 1 bar with the Nosè–Hoover thermostat[58,59] and Parrinello–Rahman barostat.[60] The thermostat relaxation constant was chosen to be

2.0 ps, and the barostat relaxation constant was 5.0 ps. The α-helix, β-strand, PPII, and the unrestrained groups were sampled both at 310 K and a slightly elevated temperature of 360 K. 310 K is close to the human body temperature, while the 360 K trajectories enhance the sampling of the conformation space. The survey group (d) was only simulated at 310 K.

Sixty conformations were collected for each of the five groups mentioned above. For the survey group (d), 15 conformations were saved for each of the four subsets. For the other groups, 30 conformations were saved from each of the two temperatures. The total number of conformations per generation is thus 300. The five groups of conformations are summarized in Table 1.

Table 1. Torsional Constraints for Different Conformation Groups^a

group	Φ (deg)	Ψ (deg)	T (K)	N
α-helix^b	-60	-45	310, 360	30, 30
β-sheet	-135	-135	310, 360	30, 30
PPII	-75	150	310, 360	30, 30
survey group	-120	N/A	310	15
	-150	-60	310	15
	-150	60	310	15
	-90	60	310	15
unrestrained	N/A	N/A	310, 360	30, 30

^{a}Simulation temperatures (T) and number of frames (N) used are also reported. ^{b}The α-helix group was sampled with additional restraints of H–O distance and N–H–O angle (see text). N/A: not applicable since no constraints were made.

2.2. QM/MM Step. The reference forces for AFM were obtained in the QM/MM step. For each conformation in the training set, the following procedure was used to identify a QM region and the associated MM environment for Coulombic embedding.

(a) Ala$_7$ is included in the QM region.

(b) If a water molecule has any atom within 4.5 Å of a carbon atom or within 3.8 Å of any other atom, it will be included in the QM region. 4.5 and 3.8 Å are approximately the location of the first minima in the corresponding peptide and water RDFs.

(c) Randomly select five water molecules from those identified in step (b). All water molecules within 2.6 Å of the selected water molecules are also included in the QM region.

(d) Any water molecule within 7 Å of any heavy atoms of the alanine but not included in the QM region are identified as an MM water. Water molecules further away will be omitted.

(e) The Ala$_7$ and any water molecule not having any MM atoms within 2.6 Å will be fitted. The remaining QM waters will not be fitted but are still treated quantum mechanically to shield the fitted atoms from nearby MM point charges.

With this procedure, the QM region will include from 58 to 100 water molecules depending on the conformation, although only 12 water molecules are fitted on average. All MM waters are represented as point charges for the QM/MM calculations.

For the QM/MM conformations, reference forces were calculated with density functional theory (DFT) using the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional with the aug-cc-pVDZ basis set for all the heavy atoms and

the cc-pVDZ basis set for hydrogen. The calculations were performed using the Parallel Quantum Solutions program with Fourier transform Coulomb for two-electron integrals.[61] The self-consistent field calculations were performed with a Brillouin convergence of 10^{-6}. The zero tolerance for one-electron integrals was 10^{-11} and that for two-electron integrals was 10^{-10}.

2.3. FM Step. In the FM step, the force fields were determined by fitting the QM/MM reference forces. Due to the large number of parameters in the alanine potential, for each generation of FM, reference forces from up to two previous generations of AFM are combined and used as a training set. This procedure will use up to 900 configurations to be fitted in each generation.

Figure 1 summarizes the 14 atom types to be used, which were assigned based on the chemical environment of each atom. The parameters were determined with a three-step fitting procedure.
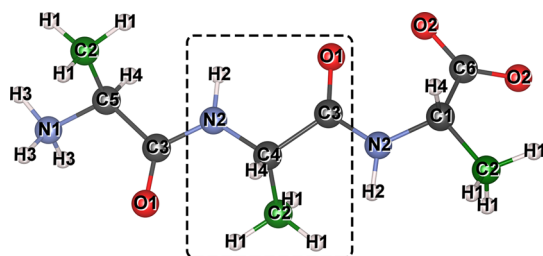

Figure 1. Fourteen atom types used for the development of the poly-alanine peptide force field. The dashed box indicates the residue that repeats for longer peptides.

2.3.1. Determination of Dispersion Parameters. The C6 dispersion terms will be fitted to Grimme's D3 dispersion.[62,63] In order to reduce the number of interaction terms for more efficient MD simulations, dispersion terms were only placed between heavy atoms. Three peptide fragments (Figure S1), representing the alanine residue, the C terminus, and the N terminus, along with isolated water molecules were used to fit the dispersion parameters.

D3 computations were performed on conformations with two isolated peptide fragments and two water molecules. The conformations were extracted from a hydrated Ala$_7$ MD simulation. The peptide fragments extracted are required to be separated by at least 5 Å in nearest atom distances. The two randomly selected water molecules are required to be between 4.5 and 6.5 Å from any one of the two peptide fragments and at least 5 Å from each other. The distance criterion was applied to ensure that only the long-range asymptotes were fit. Very distant pairs were also avoided to ensure good numerical stability.

The D3 forces were computed with the Becke−Johnson (BJ) damping function.[64−66] Since the reference forces were computed with PBE, parameters for the D3 computations were chosen to be compatible with the functional.

Since no combination rules were being used, an independent C6 parameter is fitted for each pair of unique atoms. For the fitting of C6 terms, the force and torque on each peptide fragment were used as reference. Since only a single dispersion site was used for water, only the total force of water was fitted in this step.

In our force field, the dispersion terms will be modeled with the short-range damped (SRD) form with the expression

$$V_{\text{disp}}(r_{ij}) = \frac{C_{6ij}}{r_{ij}^{6} + r_0^{6}} \qquad (1)$$

where r_0 is 0.6 times the sum of the atomic van der Waals (vdW) radius[67] summarized in Table S1 in the Supporting Information. A comparison of the SRD dispersion and the popular Tang−Tonnies dispersion is also presented in the Supporting Information. The SRD form provides much stronger damping inside the vdW radius and is thus more stable. Also, the SRD form consistently results in lower root-mean-square error of the fit for the final AFM force field when compared to Tang−Tonnies.

We note that since the Grimme D3 dispersion is empirical, it would be possible to simply use the D3 dispersion as is. However, such an approach would require modification of the Gromacs code. Our goal is for our final force field to be compatible with the majority of the MD packages available in the public domain. In addition, recasting the D3 dispersion, which is between every pair of atoms, to a heavy atom only description reduces the computational cost, which is meritorious for a biological force field. The only problem with such a remapping is that certain fitted dispersion parameters can become positive. Such repulsive dispersion is unphysical and indicates overfitting. To address this problem, the dispersion terms were refit after the offending terms are removed sequentially. The final force field only retained 46 dispersion terms total.

2.3.2. Determination of Partial Charges and the Intermolecular Short-Range Nonbonded Parameters. In step (b), the parameters of the dispersion terms will be fixed and Coulombic terms and the peptide-water short-range nonbonded terms are fitted. The reference forces will be PBE forces computed in the QM/MM step and corrected with D3(BJ) dispersion.

When fitting charges, charge constraints have been applied so that the −C=O−NH−Cα−H and the side-chain −CH$_3$ group are both neutral. The sum of the charges for the N-terminus O=C−[CH]−[NH$_3^+$] will be 0.766 e, and the [NH]−[CH]−[CO$_2^-$] group at the C terminus will be −0.766 e. The use of 0.766 e for the excess charge is based on previous work of fitting monovalent ions with AFM.[53] Such a charge is expected to compensate for the implicit treatment of optical contribution to the dielectric constant.[53,68]

The short-range repulsion terms will be modeled with an exponential function

$$V_{\text{rep}}(r_{ij}) = A_{ij}e^{-\alpha_{ij}r_{ij}} \qquad (2)$$

The repulsion terms are placed between every peptide heavy atom and water oxygen. In addition, repulsion terms are placed between a hydrogen atom and any atoms with a negative charge. For peptide−water repulsions, α_{ij} in eq 2 was fitted as a nonlinear parameter.

2.3.3. Determination of Intramolecular Parameters. After determining the intermolecular terms, intramolecular terms were fitted with all intermolecular parameters fixed to the value determined in steps (a) and (b). The alanine force field contains bond, angle, and dihedral terms and the Buckingham potential for the intramolecular short-range nonbonded interactions. For the Buckingham potential, only the exponential repulsion is fitted in this step since the dispersion terms were determined in step (a).

The assignment of bond and angle terms is rather straightforward. Every covalent bond will be modeled with a harmonic bond potential. Any two bonds sharing an atom will have a harmonic angle term. We note that for any sp^3 atomic center with four different bonds, six different angle parameters can be fit. However, there are only 5 degrees of freedom for the angles; thus, the six angle terms will be dependent on each other. For a planar sp^2 atom, technically only two free angles exist. However, we allow all the three angles to be fit. The sum of the three equilibrium angles can be larger than 360°. Since a pyramidal vertex has the sum of the three angles less than 360°, the larger sum provides a driving force to keep the center planar and is thus desirable.

Torsional interactions for the peptide will be described using the cosine dihedral formula

$$V_{dih}(\theta) = k_n(1 + \cos(n\theta - \delta)) \quad (3)$$

where θ is the torsional angle.

The following three types of dihedral terms are used:

i. The single-bond potential (SP) has a multiplicity n of 3 and the equilibrium angle, δ, will be fitted. The fitting of δ allows the minimum of the torsional potential to differ from 60°; however, care must be taken for chiral atoms when δ is different from zero.

ii. The double-bond potential (DP) has a multiplicity n of 2 and a δ of 180°.

iii. A triple-term potential (TP) is composed of three cosine dihedral terms with n of 1, 2, and 3. δ will be zero for TP to avoid overfitting with too many parameters to describe one torsional angle.

In order to avoid overfitting, a minimal set of dihedral terms is included. Each dihedral is defined with a four-atom sequence. Counting from the N-terminus, each unique atom type will be the first atom of a four-atom sequence. The three other atoms will use as many backbone atoms as possible.

Since the peptide bond ω dihedral has a double-bond character, it will be described with DP dihedral. All atoms bonded to the C3−N2 will have a DP dihedral to help the atomic centers to remain planar. The only torsional angles that use the TP bond are the peptide ϕ and ψ angles. To better define the location of the side-chain methyl group, a separate torsional term is fitted to describe the ϕ' and ψ' dihedrals. However, for ϕ' and ψ' dihedrals, only the SP term is used. Table 2 provides a summary of all the torsional terms fit.

As with most force fields, no nonbonded short-range interactions will be used when two atoms are separated by one or two covalent bonds. 1−4 interactions are interactions between atoms separated by three bonds. 1−4 interactions are sometimes scaled in other force fields.[17,18] To avoid introducing another parameter as the scaling factor, we decided to include 1−4 interactions without scaling.

It is worth mentioning that fitting without 1−4 interactions was also attempted. Ignoring 1−4 interactions would lead to problematic distributions. For example, in what is known as the C_S conformation for the alanine peptide,[69,70] the O1−H2 atoms (see Figure 1) will be close. The O1−H2 interaction is 1−5; however, the H2 does not see the Coulombic repulsion from C3 without 1−4, leading to problems that are difficult to fully compensate by torsional terms alone.

All atom pairs could have intrapeptide repulsion as long as they are separated by more than two bonds. Additional challenges arise in the fitting of such repulsion since bonded

Table 2. Torsional Terms in the Force Field[a]

	torsional terms	fitting type
ϕ	C3−N2−C4−C3	TP
	C3−N2−C1−C6	
ψ	N2−C4−C3−N2	TP
	N1−C5−C3−N2	
ϕ'	C3−N2−C4−C2	SP
	C3−N2−C1−C2	
ψ'	C2−C4−C3−N2	SP
	C2−C5−C3−N2	
ω	C4−C3−N2−C4	DP
	C5−C3−N2−C4	
	C4−C3−N2−C1	
	O1−C3−N2−H2	
	C5−C3−N2−H2	
	C4−C3−N2−H2	
	O1−C3−N2−C4	
	O1−C3−N2−C1	
other	H1−C2−C5−C3	SP
	H1−C2−C4−C3	
	H1−C2−C1−N2	
	H4−C4−C3−O1	
	H3−N1−C5−C3	

[a]See text for discussion of various fitting types.

terms limit the range of possible interatomic distances sampled in the training set. Based on interatomic distances sampled in the training set, we classify intrapeptide atom pairs into distant, near, and intimate pairs. Distant pairs are atom pairs that are far apart and do not feel steric repulsion even at the shortest distance in the training set. This is defined as atom pairs that never get closer than 3.5 Å for two heavy atoms or 3.0 Å if at least one of the atoms is hydrogen. Near pairs have distances shorter than that of the distant pair but greater than 2.0 Å. Atom pairs that get closer than 2.0 Å will be referred to as intimate pairs.

For distant pairs, the steric repulsion is never felt in the conformations in the training set. Thus, no sufficient information exists from the QM/MM calculations to fit repulsion. For these pairs, the Amber ff99SB $1/r^{12}$ repulsion will be used between two heavy atoms. Borrowing of the Amber repulsion for distant pairs will not lead to inferior quality of the force field since those repulsions do not contribute significantly at the temperature of interest. The use of a trajectory with elevated temperature of 360 K for our training set generation was designed to improve sampling of steric encounters to better capture cases where an AFM repulsion would be important.

For near pairs, α_{ij} in eq 2 was deduced from fitting to SAPT exchange repulsion between similar functional groups. In such a case, a scan was performed using SAPT with the exchange repulsion energies from the scan fit to eq 2, α_{ij} determined from the fit will be used as is for AFM. For pairs where such a scan cannot be easily performed without leading to a steric clash of other atoms, α_{ij} was fixed to 3.6 Å^{-1} for repulsion between heavy atoms, and to 3.5 Å^{-1} if at least one of the two atoms is hydrogen. Such a SAPT-based predetermination of α_{ij} reduces the number of nonlinear parameters that needs to be fitted. For intimate pairs, α_{ij} is fully optimized with AFM.

We note that occasionally the fitted exponential term turns out to be attractive. This is more likely to happen between pairs of atoms already with strong Coulombic repulsion. In such a case, we will remove the attractive exponential term and refit to ensure all exponential terms to be repulsive.

After the FM step, the new force field will be used to start a new generation of AFM. The final force field was determined by a global fit including all conformations in the final six generations of AFM. The force field parameters are summarized in the Supporting Information along with a set of GROMACS input files that can be used to simulate zwitterionic Ala$_7$ in water.

3. COMPUTATIONAL DETAILS FOR MD SIMULATIONS

All the force field fitting was performed with the CRYOFF code version 2.7.2. The code optimizes linear parameters with singular value decomposition and nonlinear parameters with the simplex method. The AFM workflow was performed using AFM toolkit 1.1. In addition, all MD simulations were performed with the GROMACS simulation package.[71]

The J-coupling constants $^3J(H_N,H_\alpha)$, $^3J(H_N,C')$, $^3J(H_\alpha,C')$, $^3J(H_N,C_\beta)$, $^1J(N,C_\alpha)$, and $^2J(N,C_\alpha)$ were calculated with the Karplus equations[48]

$$J(\varphi) = A \cos^2(\varphi + \theta) + B \cos(\varphi + \theta) + C \tag{4}$$

where φ is the backbone dihedral angle. The coefficients A, B, and C and the offset angle θ used were the same as those used by Graf et al.[40,72−74] and summarized in the Supporting Information. $^3J(H_N,C_\alpha)$ was calculated with the following relation[75]

$$
\begin{aligned}
^3J_{HNC_\alpha}(\phi_i, \psi_{i-1}) = {} & -0.23 \cos \phi_i - 0.20 \cos \psi_{i-1} \\
& + 0.07 \sin \phi + 0.08 \sin \psi_{i-1} \\
& + 0.07 \cos \phi \cos \psi_{i-1} \\
& + 0.12 \cos \phi \sin \psi_{i-1} \\
& - 0.08 \sin \phi \cos \psi_{i-1} \\
& - 0.14 \sin \phi \sin \psi_{i-1} + 0.54
\end{aligned}
\tag{5}
$$

The quality of the agreement is presented as[76]

$$\chi^2 = \frac{1}{N} \sum_i^N \frac{(\langle J_i \rangle_{sim} - J_{i,expt})^2}{\sigma_i^2} \tag{6}$$

where σ_i is the systematic error for the couplings determined from the Karplus equation. The values of σ_i used in this work are summarized in the Supporting Information (Table S2). We computed χ^2 including the seven constants, $^3J(H_N,H_\alpha)$, $^3J(H_N,C')$, $^3J(H_\alpha,C')$, $^3J(H_N,C_\beta)$, $^1J(N,C_\alpha)$, $^2J(N,C_\alpha)$, and $^3J(H_N,C_\alpha)$. In addition, we also computed χ^2 leaving out $^3J(H_N,C_\alpha)$ as will be discussed in section 4. In our work, J-coupling constants related with the ψ dihedral of the N terminus of the peptide and the ϕ dihedral of the C terminus were not included when evaluating χ^2. Different studies may include a slightly different set of J-coupling constants and have used different σ values when evaluating χ^2. Such differences lead to complications when comparing with literature values. Unless otherwise noted, the χ^2 reported in this work maintains consistent use of σ_i and number of terms to ensure a fair comparison. For the fit to experiments by Graf, the χ^2 value was recomputed with fitted J-coupling constants in the literature[40] but with our choice of σ_i and number of terms. For other force field models, all J-coupling constants were recomputed with our simulations under the same condition as the AFM2020 model.

Liquid-phase simulations were performed in cubic boxes with dimensions of 2.7, 3.5, and 4.1 nm, respectively, for zwitterionic Ala$_3$, Ala$_5$, and Ala$_7$. A dodecahedron box was used for Ala$_{19}$ to reduce the number of water molecules needed. The Parrinello−Rahman barostat and Nosè−Hoover thermostat were used to maintain pressure to 1 bar and temperature to 300 K for all liquid-phase simulations.

Since the force field was fitted only to short peptides in solutions, it is important to assess how the force field will perform in the absence of hydration water. To accomplish this assessment, gas-phase simulations were performed to study the AFM2020 model without hydration water. We note that these simulations do not necessarily reflect the true behavior of the peptide in the gas phase, which was never fitted in the development of the model. The gas-phase simulation is only used to gauge the performance of the peptide models in the absence of competing hydrogen bonds from the solvent and in a small dielectric medium, such as the interior of a protein. Without a high dielectric solvent, simulating zwitterionic peptides in gas phase leads to unphysical attractions between charged termini. To address this issue, gas-phase simulations were performed with Ace-(Ala)$_n$-NMe. The atom types for terminal C=O, N−H, and CH$_3$ are the same as those used for the solution-phase alanine. However, the charges on CH$_3$ were slightly adjusted to make Ace and NMe neutral. The adjusted charges and the extra bonds and angle parameters are shown in the Supporting Information. Stochastic rescaling thermostats[77] were used for all gas-phase simulations. A set of GROMACS input files for simulating the gas-phase Ace-(Ala)$_7$-NMe are also provided in the Supporting Information.

Classifications into different types of secondary structure are based on the range of torsional angles as shown in Table 3.

Table 3. Definition of Different Regions of Conformational States

region	torsional range
α	$-160° < \phi < -20°, -120° < \psi < 50°$
β	$180° < \phi < -90°, 50° < \psi < 240°, 160° < \phi < 180°, 50° < \psi < 180°$
PPII	$-90° < \phi < -20°, 50° < \psi < 240°$

4. QUALITY ASSESSMENT OF THE AFM FORCE FIELD

In this work, we will refer to the alanine force field developed AFM2020. The χ^2 values of AFM2020 were calculated for zwitterionic Ala$_3$, Ala$_5$, and Ala$_7$. Since a potential for counterion has not been developed, zwitterionic simulations best resemble the cationic form measured experimentally.

Table 4 reports the χ^2 computed with the seven J-coupling constants summarized in Section 3 and labeled 7J. The simulations for AFM2020 and the other force fields were performed for 1 us each with the same zwitterionic form. Figure S3 in the Supporting Information shows the running average of χ^2 and the relative free energies for two separate 500 ns simulations of Ala$_7$ using the AFM2020 model. χ^2 converged in less than 50 ns in each case, indicating that sufficient sampling has been accomplished for this model. The relative free energies of the two 500 ns simulations also agree well, indicating that the 1 us total simulation time is adequate. Shorter peptides are expected to take less time to reach convergence. It is clear from Table 4 that the AFM2020 model, without any empirical parameters, performs better than Amber ff99SB,[21] CHARMM27/CMAP (C27/CMAP),[13,78] Graf's thermodynamic fit,[40] and CHARMM36m (C36m).[26]

Table 4. χ^2 Values for Ala$_3$, Ala$_5$, and Ala$_7$ with Different Force Fields Calculated for the 7J, 6J, and 7J* Definitions of χ^2 (See the Text)

	7J			6J			7J*		
	Ala$_3$	Ala$_5$	Ala$_7$	Ala$_3$	Ala$_5$	Ala$_7$	Ala$_3$	Ala$_5$	Ala$_7$
Amber ff99SB	3.34	3.00	2.49	3.34	2.56	2.43	0.67	0.49	0.46
C27/CMAP	3.24	3.65	3.03	3.24	2.77	2.72	0.64	0.50	0.48
C36m	1.28	0.99	0.71	1.28	0.54	0.60	0.18	0.07	0.07
Graf[40]	0.57	1.01	0.80	0.57	0.55	0.76	0.08	0.17	0.17
AFM2020	0.45	0.86	0.40	0.45	0.31	0.25	0.07	0.05	0.03

We note that while C36m published in 2017 was fairly new, no recent models of the Amber family were tested in our work. There are multiple lineages for the Amber models. Some models, such as Amber14sb, seem to reduce the χ^2 by a factor of 2.[22] The IPQ lineage models reduce the χ^2 even further.[15,16] We tested the ff99SB model because it has been studied extensively by many groups, such as the early work of Best et al.[76] Reproducing ff99SB J-coupling in the literature provides a good consistency check for our study in terms of simulation protocol and time to convergence for peptides of various lengths.

The 7J χ^2 computed by Amber ff99SB is larger than that of 1.8 reported previously in the literature.[76] However, this is due to the use of different σ. With the σ increased by 30% as had been done in the literature, our simulation produced an ff99SB χ^2 of 2.1, which is in good agreement with the literature value. It is worth pointing out that the experimental J coupling was measured on a cationic system in a buffer solution, as did some of the previously published simulations. Because of the lack of counterion parameters currently in AFM2020, our simulations with AFM2020 and the other fields were all performed with zwitterionic peptide in neat water. It is our assumption that the effect of the minor difference between simulated and experimental conditions on peptide conformation is small. This assumption is supported by previous simulations reported in the Supporting Information of the work of Best et al.[76]

We note that during the parameterization of the C36m model, the Graf J-coupling constants for Ala$_5$ were part of the training goals. In other words, C36m was fit directly to produce the scalar coupling data for Ala$_5$ directly. It is worth mentioning that C36m tried to reach a compromise with many other data being fit and did not make reducing χ^2 its sole objective. It is interesting to note that our model performs slightly better than the C36m model for Ala$_5$ and significantly better than for Ala$_3$ and Ala$_7$.

We find it especially encouraging that the AFM2020 model performs better than the Graf thermodynamic fit. The Graf thermodynamic fit was produced by fitting the population of each conformation to minimize the deviation between calculated and measured J-coupling constants. In Graf's fit, the σ_i value of each J coupling is not considered. The relative population from Graf is not predicted by any force field, and only the J-coupling within each conformational state was predicted by a force field. The Graf work studied several force fields exploring relative performance and chose the GROMOS force field to perform J-coupling estimates for the population fit. Our model performs better than Graf's direct thermodynamic fit in all cases especially for Ala$_7$. This indicates that our model not only gets the relative population of different secondary structures in good agreement with experiments but also better reproduces the finer scale conformational distributions within each basin.

Although for Ala$_5$ χ^2 from AFM2020 is better than both C36m and Graf's thermodynamic fit, it is still worth additional investigation since AFM2020 seems to have relatively the

worst performance for this peptide when compared to Ala$_3$ or Ala$_7$. A careful examination of the contribution to χ^2 indicates the $^3J(H_N,C_\alpha)$ term being the largest single contribution to χ^2. Ala$_3$ does not have experimental $^3J(H_N,C_\alpha)$ as a reference and is thus not affected. Both Ala$_5$ and Ala$_7$ have two experimental $^3J(H_N,C_\alpha)$ values available. However, the χ^2 for Ala$_7$ has 31 terms while that for Ala$_5$ only has 18 terms. Thus, the large contribution from $^3J(H_N,C_\alpha)$ adversely affects the χ^2 of Ala$_5$ the most. The large contribution from this term is due to the small σ^2 of 0.01 associated with it.

To confirm this hypothesis, we compute both the χ^2, where σ_i in eq 6 is dropped, and the χ^2, where the $^3J(H_N,C_\alpha)$ contribution is eliminated. The former is identical to the objective function used by Graf and is labeled as 7J*. The latter is labeled 6J in Table 4. To ensure a fair comparison, the χ^2 values of all the models were recomputed with the same definitions. It is clear from Table 4 that AFM2020 performs better than all other models by a larger margin percentage-wise for all alanine lengths studied in this work when the 6J or 7J* χ^2 is used as the figure of merit. For 7J*, where the Graf study fitted the conformation of each residue to optimize, our χ^2 is better by a factor of 3 for both Ala$_5$ and Ala$_7$.

Table 5 reports each individual J-coupling constant for Ala$_7$, which allows for a detailed residue by residue comparison. For $^3J(H_N,H_\alpha)$, which is the most sensitive indicator of the ϕ angle, experimental data show that $^3J(H_N,H_\alpha)$ increases from the N terminus to C terminus, indicating a transition from more PPII conformations to more β conformations along the chain.[40] The same trend is also captured by the simulated $^3J(H_N,H_\alpha)$. The most sensitive indicator for the ψ angle, $^2J(H_N,C\alpha)$, decreases along the chain from N-terminus to C-terminus. This trend is interpreted as an increase of helical content along the chain[40] and is also captured by the AFM2020 model. It is encouraging that not only do the J-coupling constants produced by the AFM2020 model give a small χ^2, but the AFM2020 peptide also captures key relative trends observed experimentally. The predictions based on PBE/D3 are thus in quantitative agreement with experiments as far as the scalar couplings are concerned. Detailed prediction for each individual J-coupling for Ala$_3$ and Ala$_5$ is reported in the Supporting Information and will not be elaborated here.

Table 6 reports the population of each conformational state for Ala$_3$, Ala$_5$, and Ala$_7$ from AFM2020 along with that from Graf's thermodynamic fit. It is worth noting that the conformation of each residue is defined by torsional angles only according to Table 3. This is the same as Graf's definition to facilitate comparison with his population estimates. However, an α-helix according to this definition may not have an $i \rightarrow i + 4$ hydrogen bond.

While Graf's thermodynamic fit predicts no conformation in the α helix state and 80% PPII for Ala$_3$, Ala$_5$, and Ala$_7$, the AFM2020 model predicts a lower PPII population of 60−65%

Table 5. J-Coupling Constants of Ala$_7$ Calculated Using AFM2020 along with Experimental References[40]

residue	type	J-coupling constants/Hz	
		AFM2020	exp
A2	$^3J(H_N H_\alpha)$ (ϕ_2)	5.83	5.61
	$^3J(H_N C')$ (ϕ_2)	1.04	1.15
	$^3J(H_\alpha C')$ (ϕ_2)	1.60	1.89
	$^3J(H_N C_\beta)$ (ϕ_2)	2.04	2.31
	$^1J(N,C_\alpha)$ (ψ_2)	11.45	11.37
A3	$^3J(H_N H_\alpha)$ (ϕ_3)	5.88	5.66
	$^3J(H_N C')$ (ϕ_3)	1.06	1.20
	$^3J(H_\alpha C')$ (ϕ_3)	1.62	1.85
	$^3J(H_N C_\beta)$ (ϕ_3)	2.01	2.20
	$^1J(N,C_\alpha)$ (ϕ_3)	11.23	11.27
	$^2J(N,C_\alpha)$ (ψ_2)	8.46	8.52
	$^3J(H_N C_\alpha)$ (ϕ_3, ψ_2)	0.47	0.66
A4	$^3J(H_N H_\alpha)$ (ϕ_4)	6.09	5.77
	$^3J(H_N C')$ (ϕ_4)	1.03	1.20
	$^3J(H_\alpha C')$ (ϕ_4)	1.69	1.80
	$^3J(H_N C_\beta)$ (ϕ_4)	1.93	2.23
	$^1J(N,C_\alpha)$ (ψ_4)	11.20	11.22
	$^2J(N,C_\alpha)$ (ψ_3)	8.23	8.29
	$^3J(H_N C_\alpha)$ (ϕ_4, ψ_3)	0.44	0.56
A5	$^3J(H_N H_\alpha)$ (ϕ_5)	6.12	5.92
	$^3J(H_N C')$ (ϕ_5)	1.02	1.19
	$^3J(H_\alpha C')$ (ϕ_5)	1.70	1.56
	$^3J(H_N C_\beta)$ (ϕ_5)	1.93	2.23
	$^1J(N,C_\alpha)$ (ψ_5)	11.27	11.29
	$^2J(N,C_\alpha)$ (ψ_4)	8.19	8.22
A6	$^3J(H_N H_\alpha)$ (ϕ_6)	6.22	6.04
	$^3J(H_N C')$ (ϕ_6)	0.99	1.10
	$^3J(H_\alpha C')$ (ϕ_6)	1.72	1.67
	$^3J(H_N C_\beta)$ (ϕ_6)	1.91	2.21
	$^1J(N,C_\alpha)$ (ψ_6)	11.21	11.29
	$^2J(N,C_\alpha)$ (ψ_5)	8.28	8.24

Table 6. Population of Different Secondary Structure Motifs for Ala$_3$, Ala$_5$, and Ala$_7$ from Simulations

	AFM2020			Graf's fit		
	Ala3	Ala5	Ala7	Ala3	Ala5	Ala7
% α	20.9	21.7	15.2	0.0	0.0	0.0
% β	17.8	18.4	20.3	8.0	16.5	15.8
% PPII	61.2	59.8	64.3	92.0	83.5	84.2
% others	0.1	0.1	0.2	0.0	0.0	0.0

and about 15−20% conformations in the helix state. It is worth emphasizing that Graf's analysis is based on the best fit to experimental J-coupling. Our predicted J-coupling is in better agreement with experiments than Graf's fit. Thus, based on NMR scalar coupling constants, our study suggests that there is not sufficient evidence to rule out helix populations of up to 15% for short peptides, such as Ala$_7$.

It will be discussed later that our hydrated Ala$_{19}$ simulation shows a greater population of helix when compared to shorter peptides. It is somewhat unexpected that Ala$_7$ has around 15% population in the helix region, whereas Ala$_3$ and Ala$_5$ have almost 20% helical conformations. Since our simulation shows good agreement between populations determined from two 500 ns trajectories (see the Supporting Information), we suspect that

the difference is more than not fully converged population distributions of the peptides.

We note that it would be impossible for Ala$_3$ to form $i \rightarrow i + 4$ hydrogen bonds typical for α-helix. The helix population for all the peptides are classified solely based on the torsional angle distributions shown in Table 3. One possible explanation is that very short peptides, such as Ala$_3$ and Ala$_5$, are more flexible than Ala$_7$ and spent less time in the most preferred conformation. While Ala$_7$ is more enriched in PPII, Ala$_3$ and Ala$_5$ have smaller free energy differences between different conformations and are thus of less preference in any particular conformation. Consequently, the number of conformations end up in the α helix basin increases slightly.

Figure 2 compares the relative free energies of different conformations in the Ramachandran plot for Ala$_7$. Similar free energy profiles for C27/CMAP, C36m, and Amber ff99SB are shown for comparison. The free energy profile of AFM2020 is most similar to C36m, especially with respect to the relative free energy between β region and PPII region. AFM2020 does sample a small left-turn helix region, although such conformations were not in the training set. For the more populated regions at negative ϕ, AFM2020 provides a free energy profile quite similar to the existing models. The AFM2020 prediction could be more accurate, considering the better agreement with experimental J-coupling data. For the less populated positive φ region, the performance of AFM2020 should be taken with some caution since the current fitting has not included left-turn helix in the training set. However, no major problems can be revealed from the Ramachandran plot.

AFM2020, as well as many prior models, has shown that short alanine in solution is abundant with PPII conformations. On the other hand, alanine is rich in the α-helix region of proteins and is known as one of the best α-helix formers. It has been argued that competition of water−peptide hydrogen bond destabilizes peptide−peptide hydrogen bonds and a shielding of hydration water by hydrophobic residues significantly increases the tendency for α-helix formation.[79−81] In order to study the tendency for α-helix formation in a more hydrophobic environment, we simulated alanine polypeptide in vacuum, thus eliminating competition from hydration water. The interior of a protein has a low dielectric constant; thus, a vacuum simulation might be more similar to a peptide fragment in a protein than a solution-phase simulation.

Table 7 reports the conformation distribution for Ace-(Ala)$_n$-NMe with various numbers of residues in vacuum. It is worth reemphasizing that the AFM2020 gas-phase conformations probably will not represent true peptide conformations in vacuum. The gas-phase simulation was performed to assess the performance of the force field outside of the aqueous environment of the parameterization and to check whether the AFM2020 model will perform similarly to other protein force fields in a more protein-like environment. We followed the same definition of conformational basins as shown in Table 3. We report in Table 7 also the relative populations of different secondary structures in vacuum for other force fields that are known to be sufficiently accurate also for proteins.

Clearly, for the longer peptides, such as Ace-Ala$_7$-NMe and Ace-Ala$_{17}$-NMe, the AFM2020 model makes very similar predictions when compared to CHARMM C36m and Amber ff99SB with about 60% population in the α-helix state for Ace-Ala$_7$-NMe and 99% for Ace-Ala$_{17}$-NMe. For β state and PPII, our model is in closer agreement with ff99SB than C36m.
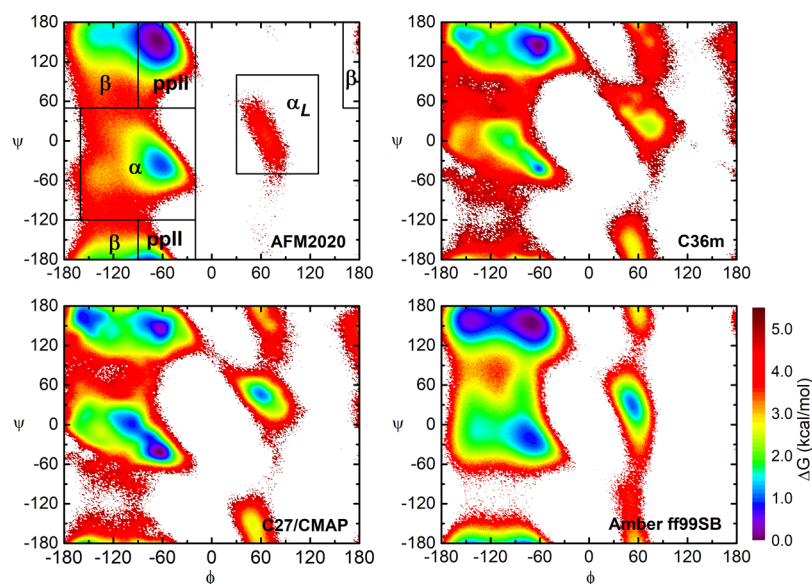
Figure 2. Relative free energies of Ala7 calculated from different force fields.

Table 7. Population of Secondary Structure Motifs of Ace-(Ala)$_n$-NMe in Vacuum from Different Force Fields

	AFM2020				C36m			
n	3	5	7	17	3	5	7	17
% α	14.1	40.0	68.1	99.6	45.6	50.8	64.9	98.7
% β	22.5	2.7	3.6	0.1	28.9	10.5	2.2	0.0
% PPII	32.6	17.7	12.9	0.2	7.1	8.7	4.1	0.0
% others	30.8	39.6	15.4	0.1	18.4	30.0	28.8	1.3
	C27/CMAP				Amber ff99SB			
n	3	5	7	17	3	5	7	17
% α	62.0	74.9	98.3	99.5	53.2	44.8	56.9	99.4
% β	21.7	5.9	0.0	0.0	12.5	4.9	2.8	0.0
% PPII	2.6	2.4	0.0	0.0	14.1	17.1	13.4	0.0
% others	13.7	16.8	1.7	0.5	20.2	33.2	26.9	0.6

The similarity of the alanine conformation distribution of AFM2020 when compared with the existing protein force fields suggests that the alanine parameters created in this work show promising prospects for adequate performance in a protein environment. Real protein simulations have to wait until a full protein force field is developed following similar procedures.

When compared to the other models in Table 7, the AFM2020 model shows the most pronounced increase of helical content as peptide length increases.[82] While the Ace-Ala$_{17}$-NMe is predominately helix for all force fields, for Ace-Ala$_3$-NMe, while AFM2020 only shows about 14% conformation in the helical state, CHARMM27/CMAP, for example, showed 62% population in the helix state for this short peptide in vacuum.

The increase of helix content for longer peptides is expected since the α-helix is anticipated to be stabilized by intrachain hydrogen bonds. For a n residue peptide, the maximum number of α-helix hydrogen bonds that can be formed is $n-4$. Thus, a shorter peptide has little stabilization from intrachain hydrogen bonds. Only for longer peptides, it would average to about one intrachain hydrogen bond per residue.

In order to further study the dependence of helix tendency on peptide length, solution-phase simulations were conducted with a zwitterionic Ala$_{19}$. The Graf work studied the HEWL-19mer, from which 30−40% α-helix content was estimated based on

NMR;[40] an estimate with circular dichroism on the same peptide gave a fractional α-helix of 8%. With AFM2020, two 350 ns simulations were performed with one from an α-helix and one from a stretched initial conformation. In each case, the first 50 ns was discarded and only the final 300 ns was used for population estimate. We only focused on residues 5−15 since they are the only residues that are expected to form two hydrogen bonds in a well-formed α-helix and we anticipate the residues closer to the termini to be less helical. For a peptide of this length, a very large error bar is expected for the mean helical population from our simulation. The two trajectories averaged to a helical content of 37 and 18%. The big difference is not surprising considering the difficulty in converging such a long peptide without utilizing an enhanced sampling technique. It is worth noting that either trajectory produced a helix content larger than that of the Ala$_7$ showing possibly a small positive correlation between helical tendency and residue length.

5. ADDITIONAL COMPARISONS BETWEEN AFM2020, AMBER FF99SB, AND CHARMM36M

It is worth noting that for the zwitterionic peptide, each terminus has a net partial charge of ±0.766 e with our model, while most other protein force fields including Amber ff99SB and CHARMM36m have this charge to be 1. The choice of 0.766 e was due to the need to compensate for the high-frequency
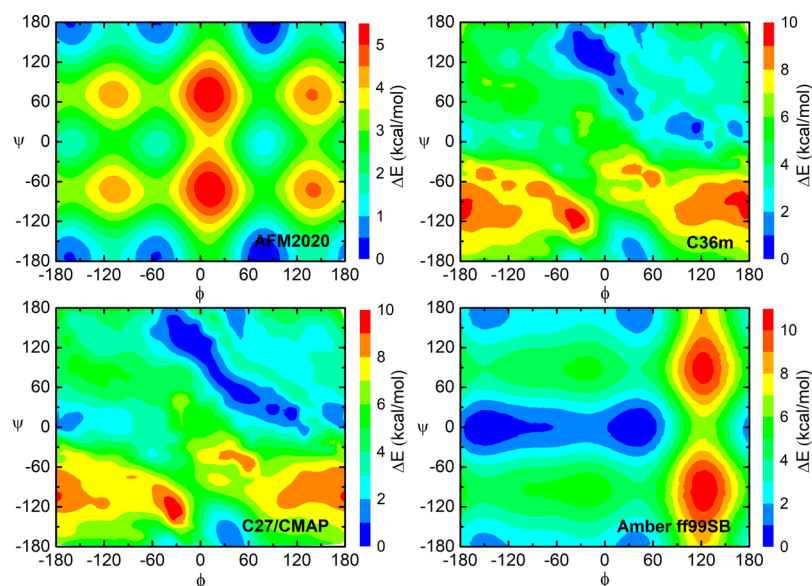
Figure 3. Contribution to the internal energy from the torsional terms only for Ala$_3$ computed with the AFM2020, CHARMM36m, CHARMM27/CMAP, and Amber ff99SB.

component of the dielectric media, which is water.[68] The charge was a fitted value taken from our previous work on ion−water potentials and was not an empirical parameter.[36,53]

Figure 3 compares contributions to the internal energy from the torsional terms of each force field plotted as a function of the ϕ and ψ dihedral. All coupled torsional terms are included, but the 1−4 nonbonded interactions are omitted for this figure. The C36m and C27/CMAP have the most features due to the use of CMAP, whereas both AFM2020 and Amber ff99SB show periodicity as a result of limiting to a linear combination of cosine dihedral terms. The scale and shape are worthy of attention since AFM2020 varies in a range of around 5 kcal/mol without particularly favoring either positive or negative side of the ϕ or ψ angles. Both C36m and ff99SB vary in a much wider range from 0 to 10 kcal/mol and favoring a particular region of ϕ or ψ angles. This observation indicates a different balance between relying on through-space effects and dihedral terms to model the torsional barrier. AFM2020 relies less on torsional terms and thus more on through-space Coulombic and steric effects with 5 kcal/mol torsional contribution, which is only about twice of a typical hydrogen bond energy. The other force fields rely on the torsional terms to do twice as much work. The torsional energy surface seems to be coaching the peptide away or toward certain conformations more.

When a force field relies on combination rules, the intrapeptide short-range nonbonded parameters are slaved to the peptide−water parameters; thus, the torsional terms have to be tasked to pick up more interactions. The AFM2020 model was developed without assuming combination rules; it thus achieves a different balance between through-space steric effects and torsional contributions.

Table 8 reports the root-mean-square difference (rmsd) between PBE/D3 forces and force field forces for conformations sampled with AFM2020 and Amber ff99SB at 310 K. The molecular force rmsd shows the performance of the intermolecular parameters in reproducing PBE/D3 reference molecular forces. The atomic force rmsd reflects the performance of both intermolecular and intramolecular energy expressions. The PBE/D3 forces were computed using the

Table 8. rmsd between Forces from PBE/D3 and the Force Fields and along with the Root Mean Square of the PBE/D3 Reference Forces (RMSF)[a]

sampling	AFM2020		Amber ff99SB	
force field	molecular force (kcal/mol·Å)	atomic force (kcal/mol·Å)	molecular force (kcal/mol·Å)	atomic force (kcal/mol·Å)
Amber ff99SB	36.47	22.31	22.28	23.00
C36m	22.43	31.61	25.0	34.8
AFM2020	8.43	9.64	11.6	12.1
PBE/D3 (RMSF)	25.34	37.25	44.39	42.63

[a]The configuration samplings were done with AFM2020 and Amber ff99SB.

same QM/MM set up as done in our reference force calculations for AFM. It is clear that the AFM2020 model provides a much smaller force rmsd when compared to the PBE/D3 reference than either Amber ff99SB or CHARMM36m. It is probably a good assumption that the AFM2020 model is sampling conformation space most similar to that would be sampled with AIMD based on PBE/D3 when compared to the other two force fields. With the AFM2020-sampled conformations, the Amber ff99SB molecular force rmsd is particularly large being 36.5 kcal/(mol·Å), likely indicating that Amber ff99SB overestimates certain intermolecular distances relative to PBE/D3. At the intermolecular distances sampled by AFM2020 or PBE/D3, the ff99SB thus predicts strong steric repulsions, leading to large rmsd from the reference forces. When ff99SB is used for sampling, the rmsd between AFM2020 and the DFT/D3 reference forces increases. This is expected since AFM2020 is not optimized specifically for such conformations. It is not surprising that even for the conformations sampled by Amber ff99SB, AFM2020 still predicts forces in much better agreement with PBE/D3 reference than either Amber ff99SB or CHARMM36m.

The last row of Table 8 reports the root mean square of the reference forces (RMSF). The RMSF will be greater than zero since the configurations were sampled at a finite temperature

and were not at minimum energy geometries. When the conformations are out-of-equilibrium, the RMSF will be more likely to increase. The molecular force RMSF is computed by the sums of atomic forces of each molecule and thus does not have contributions from intramolecular terms. The larger 44 kcal/(mol·Å) molecular force RMSF indicates that the conformations sampled by Amber ff99SB would correspond to slightly out-of-equilibrium configurations on the DFT/D3 potential energy surface. On the other hand, the minor increase of atomic RMSF upon sampling with Amber ff99SB as compared to AFM2020 indicates that the Amber ff99SB intramolecular surface is of relatively better quality when compared to the ff99SB intermolecular surface.

6. SUMMARY AND CONCLUSIONS

With AFM, a force field for hydrated zwitterionic poly-alanine has been developed based on DFT reference forces computed with the PBE exchange−correlation functional with the D3 correction for dispersion. Without fitting to any experimental data, the new AFM2020 force field provides quantitative agreement with experimental J-coupling constants.

For a hydrated alanine peptide, the χ^2 of AFM2020 is better than many existing force fields, such as Amber ff99SB and CHARMM36m. The predicted J-coupling values capture important trends from the N terminus to the C terminus consistent with experimental observations. When compared with the thermodynamic fit where the population of each conformational state is fully adjusted to minimize the deviation between calculated and experimental J-coupling constants, AFM2020 was able to produce a lower χ^2 without any empirical adjustments. We note that our investigation compares J-coupling constants in zwitterionic simulations to experimental values of a cationic system. Although previous simulations have shown that little difference in conformations between the two environments exists,[76] more definitive comparison should be performed in the future after AFM-based counterion models are parameterized.

The AFM2020 model predicts 15% helical conformations for Ala$_7$ in aqueous solutions, along with 20% β conformations and 65% PPII. Previous thermodynamic fit to the same J-coupling data used as validation in this work indicates a virtual absence of α-helix conformations for an alanine peptide of such a length. Although it may be premature to conclude that our helical conformation estimate is more realistic, the study provides some evidence that the view of a virtual absence of conformations in the helical basin is worthy of additional discussion.

It is worth noting that the experimental J-coupling data only provide insights into the conformation distribution of short hydrated peptides, and there is no direct evidence on the performance of our AFM2020 alanine model in a more protein-like environment. Simulations were performed with our AFM2020 parameters in vacuum. In the absence of water competing for hydrogen bonds, a longer alanine peptide in vacuum is predominately helical with our model, while much shorter peptides, such as Ace-(Ala)$_3$-NMe and Ace-(Ala)$_5$-NMe, have significantly lower helical content. Overall, the Ramachandran plot of AFM2020 in vacuum is similar to those of the existing protein force fields, suggesting a good chance that the AFM-based model will produce reasonable performance in a more hydrophobic environment, such as the interior of a protein.

AFM2020 is not an incremental improvement of an established force field. It is a whole new model with all intermolecular and intramolecular terms fitted self-consistently. Without using combination rules, the AFM2020 model optimizes both peptide−peptide through-space and peptide−water nonbonded interactions independently. The systematic SVD-based procedure allows complete refitting of all parameters if any future deficiency is found and the reference conformations are updated.

This work shows the promise of using AFM to create force fields for short peptides. It is a small step toward creating a whole new generation of protein force fields. Residues with large or charged side chains could be particularly challenging. More rigorous validation in a real protein environment will not be possible until parameters for other peptides are available. One direction worthy of additional investigation in the future is the explicit modeling of φ, ψ coupling as done in modern Amber and CHARMM force fields.[23,78] Whether it is meritorious to scale 1−4 interactions is also of interest for future investigations.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c11618.

> Fragments used for D3 calculations, vdW radius used for dispersion fitting, and the comparison of SRD and Tang−Tonnies damping; parameters used in Karplus equations for this work; atom types for Ace-(Ala)$_n$-NMe; summary of AFM2020 parameters; J coupling constants for zwitterionic Ala$_3$ and Ala$_5$; running averages of χ^2 and relative free energies of two trajectories to show the convergence of the simulations; GROMACS input files for zwitterionic Ala$_7$ in water and Ace-Ala$_7$-NMe in vacuum; and all training set conformations and PBE/D3 reference forces (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Feng Wang − *Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, United States;* ⓞ orcid.org/0000-0002-2740-3534; Email: fengwang@uark.edu

Authors

Ying Yuan − *Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, United States;* ⓞ orcid.org/0000-0002-3323-6912

Zhonghua Ma − *Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.0c11618

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute Comparison of Simulated and Experimental Protein-folding Dynamics. *Nature* **2002**, *420*, 102−106.

(2) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.

(3) Beauchamp, K. A.; Ensign, D. L.; Das, R.; Pande, V. S. Quantitative Comparison of Villin Headpiece Subdomain Simulations and Triplet−triplet Energy Transfer Experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 12734−12739.

(4) Shrestha, U. R.; Juneja, P.; Zhang, Q.; Gurumoorthy, V.; Borreguero, J. M.; Urban, V.; Cheng, X.; Pingali, S. V.; Smith, J. C.; O'Neill, H. M.; et al. Generation of the Configurational Ensemble of an Intrinsically Disordered Protein from Unbiased Molecular Dynamics Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 20446−20452.

(5) Herrera-Nieto, P.; Pérez, A.; De Fabritiis, G. Characterization of Partially Ordered States in the Intrinsically Disordered N-terminal Domain of p53 Using Millisecond Molecular Dynamics Simulations. *Sci. Rep.* **2020**, *10*, 12402.

(6) Wong, CF.; McCammon, AJ. Protein Simulation and Drug Design. *Advances in Protein Chemistry*; Academic Press, 2003; Vol. *66*, pp 87−121.

(7) Ponder, JW.; Case, DA. Force Fields for Protein Simulations. *Advances in Protein Chemistry*; Academic Press, 2003; Vol. *66*, pp 27−85.

(8) Guvench, O.; MacKerell, A. D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Humana Press: Totowa, NJ, 2008; Vol. *443*, pp 63−88. DOI: 10.1007/978-1-59745-177-2_4

(9) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D., Jr. Current Status of Protein Force Fields for Molecular Dynamics Simulations. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Humana Press: New York, 2015; Vol. *1215*, pp 47−71.

(10) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129−138.

(11) Halgren, T. A. The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827−7843.

(12) Rackers, J. A.; Ponder, J. W. Classical Pauli Repulsion: An Anisotropic, Atomic Multipole Model. *J. Chem. Phys.* **2019**, *150*, 084104.

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

(14) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(15) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 4515−4534.

(16) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12*, 3926−3947.

(17) 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.

(18) 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.

(19) van Gunsteren, W. F.; Billeter, S.; Eising, A.; Hünenberger, P.; Krüger, P.; Mark, A.; Scott, W.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.

(20) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99sb Protein Force Field. *Proteins* **2010**, *78*, 1950−1958.

(21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.

(22) 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.

(23) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.

(24) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, 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.

(25) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. 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.

(26) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71−73.

(27) 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 Peptidesr. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(28) 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.

(29) Nerenberg, P. S.; Head-Gordon, T. Optimizing Protein−Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011**, *7*, 1220−1230.

(30) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-principles Calculations: The Force-matching Method. *Europhys. Lett.* **1994**, *26*, 583−588.

(31) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective Force Fields for Condensed Phase Systems From Ab Initio Molecular Dynamics Simulation: A New Method for Force-Matching. *J. Chem. Phys.* **2004**, *120*, 10896−10913.

(32) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.

(33) Elrod, M. J.; Saykally, R. J. Many-Body Effects in Intermolecular Forces. *Chem. Rev.* **1994**, *94*, 1975−1997.

(34) Akin-Ojo, O.; Song, Y.; Wang, F. Developing Ab Initio Quality Force Fields From Condensed Phase Quantum-Mechanics/Molecular-Mechanics Calculations Through the Adaptive Force Matching Method. *J. Chem. Phys.* **2008**, *129*, 064108.

(35) Akin-Ojo, O.; Wang, F. The Quest for the Best Nonpolarizable Water Model From the Adaptive Force Matching Method. *J. Comput. Chem.* **2011**, *32*, 453−462.

(36) Li, J.; Wang, F. Pairwise-Additive Force Fields for Selected Aqueous Monovalent Ions From Adaptive Force Matching. *J. Chem. Phys.* **2015**, *143*, 194505.

(37) Wang, F.; Akin-ojo, O.; Li, J.; Ma, Z.; Rogers, T. R. Creat Your Own Force Field (CRYOFF). https://wanglab.uark.edu/cryoff/wanglab_CRYOFF.html (accessed July 15, 2020).

(38) Buckingham, A. D.; Fowler, P. W. A Model for the Geometries of van der Waals Complexes. *Can. J. Chem.* **1985**, *63*, 2018−2025.

(39) 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.

(40) 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.

(41) Toal, S.; Meral, D.; Verbaro, D.; Urbanc, B.; Schweitzer-Stenner, R. 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.

(42) Spek, E. J.; Olson, C. A.; Shi, Z.; Kallenbach, N. R. Alanine Is an Intrinsic α-Helix Stabilizing Amino Acid. *J. Am. Chem. Soc.* **1999**, *121*, 5571−5572.

(43) Rohl, C. A.; Fiori, W.; Baldwin, R. L. Alanine Is Helix-Stabilizing in both Template-Nucleated and Standard Peptide Helices. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3682−3687.

(44) Lin, J. C.; Barua, B.; Andersen, N. H. The Helical Alanine Controversy: An $(Ala)_6$ Insertion Dramatically Increases Helicity. *J. Am. Chem. Soc.* **2004**, *126*, 13679−13684.

(45) Nick Pace, C.; Martin Scholtz, J. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998**, *75*, 422−427.

(46) Gnanakaran, S.; García, A. E. Helix-coil Transition of Alanine Peptides in Water: Force Field Dependence on the Folded and Unfolded Structures. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 773−782.

(47) Uversky, V. N. Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Front. Physiol.* **2019**, *7*, 10.

(48) Karplus, M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1963**, *85*, 2870−2871.

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

(50) Li, D.-W.; Brüschweiler, R. Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *J. Chem. Theory Comput.* **2011**, *7*, 1773−1782.

(51) Wang, F.; Akin-Ojo, O.; Pinnick, E.; Song, Y. Approaching Post-Hartree−Fock Quality Potential Energy Surfaces With Simple Pair-wise Expressions: Parameterising Point-charge-based Force Fields for Liquid Water Using the Adaptive Force Matching Method. *Mol. Simul.* **2011**, *37*, 591−605.

(52) Pinnick, E. R.; Erramilli, S.; Wang, F. Predicting the Melting Temperature of Ice-Ih with Only Electronic Structure Information as Input. *J. Chem. Phys.* **2012**, *137*, 014510.

(53) Li, J.; Wang, F. Accurate Prediction of the Hydration Free Energies of 20 Salts through Adaptive Force Matching and the Proper Comparison with Experimental References. *J. Phys. Chem. B* **2017**, *121*, 6637−6645.

(54) Yuan, Y.; Ma, Z.; Wang, F. Leveraging Local MP2 to Reduce Basis Set Superposition Errors: An Efficient First-Principles Based Force-Field for Carbon Dioxide. *J. Chem. Phys.* **2019**, *151*, 184501.

(55) Hu, H.; Ma, Z.; Wang, F. Chapter Two - On the Transferability of Three Water Models Developed by Adaptive Force Matching. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier, 2014; Vol. *10*, pp 25−43.

(56) Rogers, T. R.; Wang, F. Accurate MP2-based force fields predict hydration free energies for simple alkanes and alcohols in good agreement with experiments. *J. Chem. Phys.* **2020**, *153*, 244505.

(57) 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.

(58) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.

(59) Hoover, W. G. Canonicarl Dynamics: Equilibrium Phase-space Distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(60) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(61) Füsti-Molnár, L.; Pulay, P. The Fourier Transform Coulomb Method: Efficient and Accurate Calculation of the Coulomb Operator in a Gaussian Basis. *J. Chem. Phys.* **2002**, *117*, 7827−7835.

(62) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(63) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(64) Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

(65) Johnson, E. R.; Becke, A. D. A post-Hartree−Fock Model of Intermolecular Interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

(66) Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

(67) Anatole von Lilienfeld, O.; Tkatchenko, A. Two- and Three-Body Interatomic Dispersion Energy Contributions to Binding in Molecules and Solids. *J. Chem. Phys.* **2010**, *132*, 234109.

(68) Leontyev, I.; Stuchebrukhov, A. Accounting for Electronic Polarization in Non-Polarizable Force Fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2613−2626.

(69) Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schaefer, L.; Momany, F. A. Ab initio Studies of Molecular Geometries. 27. Optimized Molecular Structures and Conformational Analysis of N.alpha.-acetyl-N-methylalaninamide and Comparison with Peptide Crystal Data and Empirical Calculations. *J. Am. Chem. Soc.* **1983**, *105*, 3438−3445.

(70) Cheam, T. C.; Krimm, S. Ab initio Force Fields of Alanine Dipeptide in C_5 and C_7 Conformations. *J. Mol. Struct.: THEOCHEM* **1989**, *188*, 15−43.

(71) 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.

(72) 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.

(73) Ding, K.; Gronenborn, A. M. Protein Backbone $^1H^N−^{13}C^\alpha$ and $^{15}N−^{13}C^\alpha$ Residual Dipolar and J Couplings: New Constraints for NMR Structure Determination. *J. Am. Chem. Soc.* **2004**, *126*, 6232−6233.

(74) Wirmer, J.; Schwalbe, H. Angular Dependence of $^1J(N_i, C_{\alpha i})$ and $^2J(N_i, C_{\alpha(i−1)})$ Coupling Constants Measured in J-modulated HSQCs. *J. Biomol. NMR* **2002**, *23*, 47−55.

(75) Hennig, M.; Bermel, W.; Schwalbe, H.; Griesinger, C. Determination of ψ Torsion Angle Restraints from $^3J(C_\alpha, C_\alpha)$ and $^3J(C_\alpha, H^N)$ Coupling Constants in Proteins. *J. Am. Chem. Soc.* **2000**, *122*, 6268−6277.

(76) Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J.* **2008**, *95*, L07−L09.

(77) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(78) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., III Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein Conforma-

tional Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(79) García, A. E.; Sanbonmatsu, K. Y. α-Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bondsr. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782−2787.

(80) Vila, J.; Williams, R. L.; Grant, J. A.; Wójcik, J.; Scheraga, H. A. The Intrinsic Helix-Forming Tendency of L-Alanine. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7821−7825.

(81) Williams, L.; Kather, K.; Kemp, D. S. High Helicities of Lys-Containing, Ala-Rich Peptides Are Primarily Attributable to a Large, Context-Dependent Lys Stabilization. *J. Am. Chem. Soc.* **1998**, *120*, 11033−11043.

(82) Huang, J.; MacKerell, A. D., Jr. Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the $(AAQAA)_3$ Peptide. *Biophys. J.* **2014**, *107*, 991−997.