ff19SB: Amino-acid specific protein backbone parameters trained against quantum mechanics energy surfaces in solution

Chuan Tian[1,2], Koushik Kasavajhala[1,2], Kellon A. A. Belfon[1,2], Lauren Raguette[1,2], He Huang[1,2], Angela N. Migues[2], John Bickel[1], Yuzhang Wang[1,2], Jorge Pincay[1], Qin Wu[,3] and Carlos Simmerling[*,1,2]*

1. Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States.

2. Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States.

3. Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States.

ABSTRACT

Molecular dynamics (MD) simulations have become increasingly popular in studying the motions and functions of biomolecules. The accuracy of the simulation, however, is highly determined by the molecular mechanics (MM) force field (FF), a set of functions with adjustable parameters to compute the potential energies from atomic positions. However, the overall quality of the FF, such as our previously published ff99SB and ff14SB, can be limited by assumptions that were made years ago. In the updated model presented here (ff19SB), we have significantly improved the backbone profiles for all 20 amino acids. We fit coupled φ/ψ parameters using 2D φ/ψ conformational scans for multiple amino acids, using as reference data the entire 2D quantum mechanics (QM) energy surface. We address the polarization inconsistency during dihedral parameter fitting by using both QM and MM in aqueous solution. Finally, we examine possible dependency of the backbone fitting on side chain rotamer. To extensively validate ff19SB parameters, and to compare to results using other Amber models, we have performed a total of ~5 milliseconds MD simulations in explicit solvent. Our results show that after amino-acid specific training against QM data with solvent polarization, ff19SB not only reproduces the differences in amino acid specific Protein Data Bank (PDB) Ramachandran maps better, but also shows significantly improved capability to differentiate amino acid dependent properties such as helical propensities. We also conclude that an inherent underestimation of helicity is present in ff14SB, which is (inexactly) compensated by an increase in helical content driven by the TIP3P bias toward overly compact structures. In summary, ff19SB, when combined with a more accurate water model such as OPC, should have better predictive power for modeling sequence-

specific behavior, protein mutations, and also rational protein design. Of the explicit water models tested here, we recommend use of OPC with ff19SB.

Introduction

State-of-the-art computational methods have been able to complement experimental structural biology with information that is both interesting and difficult to obtain without computers. Recent simulation highlights are the time-resolved, atomic-detail folding of ubiquitin during a 1-millisecond MD simulation[1], or the accurate reproduction of a large set of protein-ligand binding affinities[2]. Moreover, simulations are typically used during the refinement of high resolution structures obtained using experimental data such as crystallography, NMR or cryo-electron microscopy.

However, two significant caveats apply to the hypothetical power of simulations: (1) the energy function must provide an accurate model of the underlying physics of the system, and (2) the simulation must adequately sample the important regions of the resulting energy landscape. These problems are coupled, and improving the physics model typically gains accuracy at the expense of greater computational cost, reducing the conformational diversity that can be sampled. One of the main challenges in successfully employing simulations is the need to optimize this precision/accuracy compromise based on the requirements of each research project.

All-atom molecular dynamics (MD) is likely the most widely used biomolecular simulation sampling method. These often employ simple classical energy functions (force fields, FFs) which usually have many adjustable parameters, most often obtained by fitting to data from experiments or QM. Most modern FFs have very similar functional forms, but differ

significantly in choice of model systems and source of the training data. Although using even more complex models than those discussed here (such as including explicit polarizability[3]) may improve accuracy, these gains come at the cost of computational complexity and corresponding reduction in the sampling that is usually the limiting factor in the application of biopolymer force fields.

Many approximations are made in fitting FF parameters. The FFs used for simulation of biomolecules in water tend to be relatively simple, due to the large number of atom pair interactions that contribute to the net forces. In this article we focus on the FFs associated with the Amber simulation package[4], though others tend to be very similar. Amber FFs include harmonic terms for covalent structure, such as bond stretching and angle bending. The intramolecular and intermolecular nonbonded interactions are modeled as a Lennard-Jones 12-6 potential for vdW interactions, and a simple Coulomb term for electrostatics typically using fixed partial atomic charges obtained using QM-based electrostatic potentials on intact peptides. The final and crucially important component is the dihedral (torsion) correction terms, which modify the energy of the system as a function of rotation around bonds. These bond rotations control the flexibility of the biopolymer, and different corrections can alter barrier heights as well as the relative energies of various stable rotamers, directly influencing the sampled ensembles[5].

The physical motivation for the dihedral corrections is that the rest of the FF is purely classical, and therefore lacks quantum orbital effects such as the increased energy barrier for rotation around a double bond. In practice, these corrections are used broadly to empirically optimize force fields during training, accounting for quantum effects as well as other weaknesses in the simple model, such as lack of conformation-dependent polarization that could impact electrostatic interaction profiles, or even to remedy lack of agreement with experiments. In

Amber and most other atomistic FFs, the dihedral correction is modeled as a simple truncated Fourier series with amplitudes and phases that are parameters in the FF. These parameters are optimized at the last stage in order to improve the agreement between training data and MM properties calculated without the dihedral terms. Some FFs add one or more additional empirical adjustment steps to improve agreement with experiments.

Importantly, these force fields rely on an implicit assumption that each term is independent, with no coupling between parameters for bonds, angles and dihedrals. This additivity assumption extends to the nonbonded pairs as well, and is a major source of efficiency in force field calculations. In reality, coupling exists to varying extents, and parameters for one component may depend on the conformations or chemical identities of other nearby groups. This is neglected in most current biomolecular force fields. Another important key assumption is transferability: that a FF trained on one set of molecules (typically small) will perform as well on different, perhaps much larger molecules. Transferability also applies to neglecting the coupling between parameters, since it is usually assumed that one set of parameters (for example, for rotation around a bond) will perform well for multiple conformations of neighboring groups. Since transferability is imperfect, one way to improve FF accuracy is to ensure that the training data more closely reflect the situations in which the parameters will be applied, and by implicitly accounting for any coupling with neighboring groups at least in a mean-field way. Choice of model systems is therefore crucial. Enabled by greater computer power, this has led to a trend away from fitting against QM data for small organic compounds[6] to that for larger peptides.

An important example is the protein backbone φ and ψ dihedral parameters that can alter the energy profiles for these rotations, and thus influence secondary structure preferences and loop conformations. These have been frequently revised over the years based on observations of

secondary structure biases in prior models[7]. While early FFs used capped single amino acids (dipeptides) to train the backbone, our ff99SB[5] FF used tetrapeptides[8], allowing φ and ψ parameters to be trained in a context of conformational diversity of neighboring amino acids in a longer peptide. The improvement was significant, and ff99SB has been widely adopted.

Since that time, widespread use of ff99SB exposed weakness in some amino acid side chain dihedral parameters[9], probably because they were carried over from ff99 which trained them against a limited set of energy minima for simple organic compounds[6b]. In ff14SB[10], we performed complete refitting of all side chain parameters using QM data for capped amino acids. An important update was the use of multidimensional QM conformational grid scans for every side chain, rather than fitting each rotatable bond separately. Likewise, fitting was done using both α and β peptide backbone contexts. Though it stopped short of explicit dihedral parameter coupling, this approach allowed implicit inclusion of coupling of rotational profiles to neighboring groups in a mean-field way, by fitting parameters for each bond rotational energy profile in the context of multiple conformations of neighboring groups, as was done for the backbone in ff99SB. ff14SB was a notable improvement; for example, a recent study[11] of the ability of protein MD to reproduce high resolution experimental crystal data concluded that ff14SB performed best among all force fields tested, including several older Amber variants and even the empirically tuned CHARMM C36[12].

In addition to the weaknesses in side chain dihedral parameters, some studies also noted weaknesses in ff99SB backbone preferences. Several groups focused on empirically adjusting the ff99SB backbone parameters via comparison to experimental data such as NMR scalar couplings for very short peptides[13], or amino acid helical propensities[9, 13a]. Similar to these other groups, we also included in ff14SB a small empirical adjustment to ff99SB (using TIP3P

water[10]) to improve agreement with NMR data for short alanine peptides. Empirical corrections can improve performance on training data but also can be problematic when extrapolated too far. The relative scarceness of experimental data compared to the number of parameters in the FF leaves the empirical fitting problem severely under constrained. Also, the common target of NMR J coupling data is sensitive to the choice of Karplus parameters[14], they are not equally sensitive to variations in φ and ψ, and the χ^2 values typically used to score performance[15] can be highly sensitive to small details in the energy landscape yet relatively insensitive to the large differences that are observed between force fields[16]. Fitting backbone parameters to helical propensities is also challenging; it was shown that updating side chain dihedral parameters had a substantial impact on the backbone helical tendencies of some amino acids[9], perhaps because side chain positioning details may play a role in helicity by shielding backbone hydrogen bonds[7b], or due to side chain parameter changes modulating side chain entropy changes, which may influence helix formation[17]. Thus it is possible to erroneously adjust one part of the model (such as the backbone) to improve agreement with experiment, instead of fixing the more fundamental source of the error (e.g., the side chain rotamer energies). Designing or implicitly accepting cancellation of error can lead to models with unphysical and unwanted dependence between components, where one component cannot easily be improved (or even used) without exposing the compensating weakness in another.

Another major challenge to empirical fitting against experiment is deconvolution of the solvent model from the solute FF, each of which may contribute inaccuracies that lead to deviation from experiment. This is complicated further when empirical adjustment creates dependence between solute and solvent models. Shell et al. showed that the best results for predicting small protein structures were obtained using the ff96 force field with the GB-OBC implicit solvent model,

despite each having well-established deficiencies[18]. The CHARMM C22[19] FF was trained using TIP3P water, and backbone refitting was needed to use a different water model[20]. ff14ipq[21] was developed with extensive training to TIP4P-Ew[22] in the initial stages[21], requiring refitting in ff15ipq[23] to enable use with SPC/E_b water[24]. Moreover, weaknesses are apparent in studies of systems that sample diverse ensembles such as the unfolded states of proteins, or simulations of intrinsically disordered proteins (IDPs). This may well arise because of the vast number of nearly degenerate states, and the need for much higher accuracy than what is sufficient for simulating proteins in stable native basins. Currently, the challenge to FFs seems too great; for example, a recent study of IDPs found that the simulated ensembles depended dramatically on FF, but much less so on peptide sequence[25]. Piana et al. showed that the unfolded ensembles in their successful protein folding simulations were much more compact than expected from experiments[26]. Palazzesi compared simulations to NMR data, again finding generally poor agreement regardless of FF used[27]. In these and other cases, simulated ensembles are generally too compact. Several groups attempted to address the problem empirically by re-training backbone parameters against PDB coil libraries, and flattening energy landscapes[28]. Robustelli et al. carried out extensive refitting to improve the ability of ff99SB to model IDPs while retaining the ability to simulate folded proteins.[29]

More recent IDP work has implicated overly weak water-protein interactions[14, 30], consistent with other studies showing that protein-protein association in water is too favorable regardless of force field tested[31]. Best et al developed the ff03w model, empirically increasing the water-protein dispersion interaction.[14] Piana et al. developed the TIP4P-D water model, with 50% larger dispersion energies[30b], further adjusted later[29]. Both adjustments resulted in improved match to IDP experimental data such as Rg values inferred from SAXS and FRET. Recently, the

new OPC 4-point explicit water model from Onufriev et al. was shown to better reproduce liquid water properties than most other models.[32] It also results in much less compact ensembles for IDPs, suggested by the authors to be due to increased strength of the dispersion term as compared to models such as TIP3P.[33] Such studies demonstrating that newer water models improve IDP behavior again highlight the dangers in empirically adjusting specific protein FF parameters to fix what may just be a symptom of a different problem (such as solvent model). This weakens transferability, and emphasizes the value of independent development and validation of solute and solvent models.

Despite the issues described above, current force fields clearly are good enough to have enabled many excellent biophysical simulation studies. In terms of simulating global structure of proteins of various sequences, protein force fields have improved with time[34]. Current force fields typically result in stable simulations of folded proteins, with many reports of good match to experimental solution NMR observables such as NOEs, RDCs and S^2 order parameters. More challenging are studies that attempt to predict structure from sequence[35]. A particularly impressive achievement was the successful brute-force folding of ubiquitin in MD simulations[1, 36]. We reported accurate folding via MD for 16 out of 17 diverse proteins up to 100 amino acids long[37].

Despite these successes, a growing number of studies have suggested that even after the recent updates to backbone and side chain parameters, as well as water models, significant limitations remain present in protein simulations. There is a mounting consensus that current force fields do not accurately reproduce differences between backbone preferences of different amino acids. This is especially apparent in studies where the quantitative relative energies of basins are important, such as analysis of the effect of point mutations, or studies of flexible systems with

many nearly isoenergetic minima. Pande et al. suggested 6 of 19 amino acids were outliers vs. NMR and should be re-optimized[34]. Best[9] et al. and later, we reported[38] that Amber does not accurately reproduce experimental[39] amino acid specific behavior such as helical propensities, shown in **Figure 1** for ff14SB used with TIP3P. Correlation is generally poor, with most amino acids having similar helicity in simulation except the outliers of negatively charged Asp and Glu. In principle, nonbonded interactions should account for the impact of the side chain on backbone energetics (hereafter denoted "sequence dependence"), but weaknesses in the nonbonded function may limit the accuracy in modeling the short-range interactions that are responsible for backbone-side chain coupling.

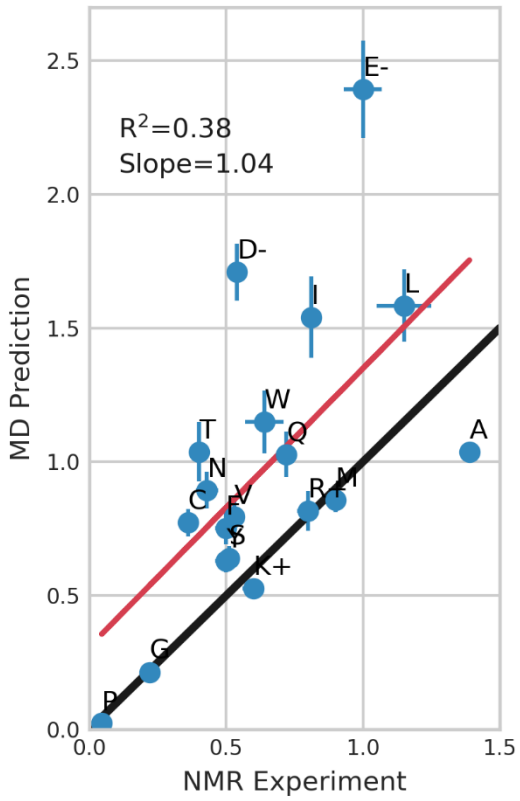

Figure 1. Helical propensities in ff14SB+TIP3P (Y) vs experiment[39] (X) for amino acids (1 letter codes). Values on the X-axis represent the data based on NMR and the reported standard

deviations.[39] Values on Y-axis represent the helical propensities fit against the combined trajectory (3.2 μs * 12), with error bars calculated via bootstrapping analysis (see **Methods: Bootstrapping analysis on helical propensity**). Black line represents perfect agreement. Linear regression (red line) was performed against the data points, with R^2 and slope quantifying the goodness of fit.

Importantly, alanine is an outlier in having its helicity significantly under predicted (below the diagonal line) while most others are over predicted. This is concerning since alanine is used as the model system in all recent Amber protein force fields for fitting of backbone dihedral parameters that also are applied to the other amino acids (except Gly). CHARMM also uses the same alanine-based backbone map for all amino acids except Pro and Gly[12, 40]. The data clearly show that empirical correction of all amino acids using alanine helicity as a target would further increase the erroneously high helical bias for the remaining amino acids.

β-branched amino acids are not modeled correctly. Experimentally, steric clash between β-branched side chains and the backbone carbonyl reduces helical propensity[39, 41]. Troublingly, simulations in ff14SB show that β-branched Ile, Val and Thr all have *higher* or similar helical propensity than Ala, the reverse of the experimental trend (**Figure 1**). In high resolution structures of folded proteins, the same trend of backbone-side chain coupling is apparent[42], where the helical basin is narrower in valine than alanine, along with a broader, flatter region at high ψ values corresponding to polyproline 2 (ppII) and β conformations as compared to alanine (**Figure 2**). It is challenging for force fields to reproduce these differences, and the alanine and valine MD Ramachandran landscapes are similar using ff14SB (see **Results**). These observations are further corroborated by solution NMR data; higher H^N-H^α scalar couplings for Val dipeptide than Ala dipeptide suggest more structures along the β-ppII transition for valine than for

alanine[43], again not reproduced in the MD data (see **Results**). The situation is similar for CHARMM C36, where errors vs. NMR remained large for valine even though the force field was empirically adjusted to obtain a good fit for alanine[45]. Taken together, the results suggest that alanine may not be an ideal model for training other amino acids, in contrast to the central assumption in >20 years of Amber and CHARMM FF development.

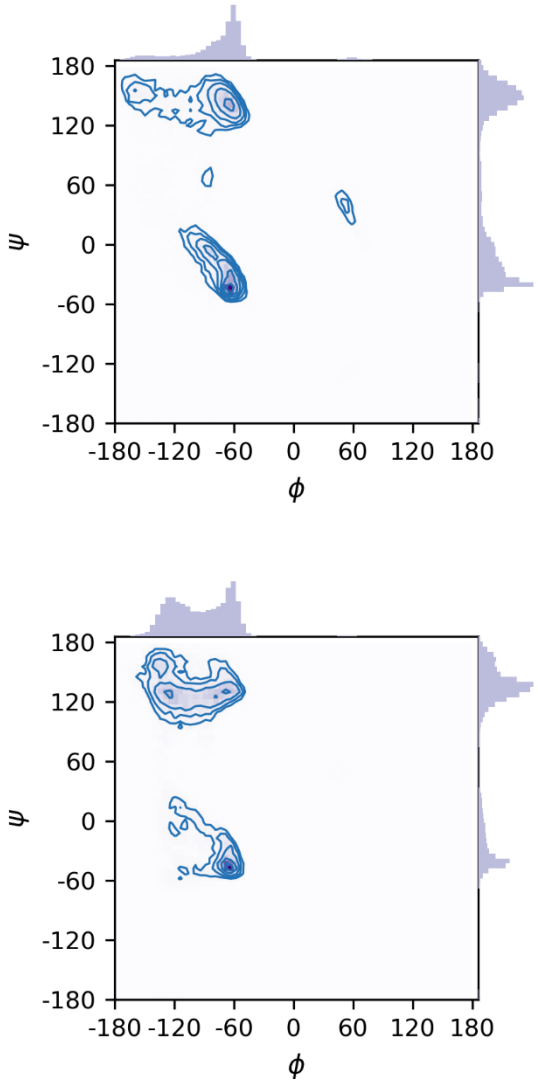

Figure 2. Ramachandran sampling in PDB shown for Ala (top) and Val (bottom) (using data from Lovell et al.[42]) Each contour line represents a doubling in population. Density is also shown

as grids filled with light (no density) to dark (maximum density). Side histograms on each subplot represent independent distributions on φ and ψ.

We previously developed empirical backbone corrections for some amino acids in order to improve residue-specific helical propensities[38]. Alternatively, Best et al. found that empirically enforcing the alanine backbone partial charges on all amino acids also resulted in improvement for charged amino acids[9], but this also may have been successful because it eliminated an inconsistency between using atom-specific partial charges and atom-type based dihedral parameters. Other recent work (for example, RSFF[28a, 44] and ff99IDPs/ff14IDPs[28b, 45]) used PDB φ/ψ distributions to develop amino-acid specific empirical backbone parameters. However, in addition to the general problems with empirical fitting discussed above, these crystal data have significant limitations that prevent them from being used as an accurate source of thermodynamic training data (such as inconsistent and cryogenic temperatures, crystal packing effects, limited or noisy data outside low-energy basins, etc.). As a specific example, although the achiral glycine should have a fully symmetric φ/ψ energy profile, PDB-based distributions show significantly enhanced incidence of glycine in the positive φ region[42], which would be reflected erroneously in force fields fit to these statistical distributions.

Going beyond empirical adjustment requires insight into the physical weaknesses in the model. What is the source of this unsatisfactory sequence dependence, despite good reproduction of QM side chain rotational energy profile data[10] in ff14SB? Speculation leads to several reasonable possibilities, including, but not limited to, lack of charge polarization of the backbone from the side chain (or weaknesses in the charge model overall), the inability of the current functional form to reproduce strong interactions between backbone and bulky side chains, or inaccurate empirical nonbonded scaling factors. Certainly using uncoupled cosine terms for backbone

dihedrals limits the accuracy attainable even with ideal QM training data or extensive empirical adjustment. The relative orientation of the two adjacent amides depends on both ϕ and ψ of the intervening amino acid, thus independent cosine terms may be insufficient at correcting the interaction energy or lack of induced polarization between these groups.

Parametrization Strategy.

In this work, we revisit the ff14SB protein backbone description with an aim to improve the performance for amino-acid specific behavior discussed above. We hypothesize that several specific weaknesses in the ff99SB strategy may be dominant factors limiting accuracy. (1) Fitting only *alanine* data, and only at the gas-phase minima, poorly constrained the resulting energy landscape for many biologically relevant conformations[46], or at locations of the slightly shifted φ/ψ minima sampled by other amino acids[42]. (2) The φ/ψ landscape is overly symmetric, arising from neglect of coupling in the simple cosine functional form. (3) Dihedral parameters are shared too broadly due to assignment by simple atom typing that does not discriminate amino acids. (4) Polarization was treated inconsistently in ff99SB and ff14SB, dating back to the original ff94 model. "Pre-polarized" Amber MM fixed partial charges[47] intended for aqueous solution simulations[48] are used while fitting dihedral parameters against gas-phase QM data, thus forcing the rotational energy profiles back towards the gas phase profiles and thereby counteracting the intended effect of better modeling charge polarization.

We describe here modifications to the protein backbone parameters that at least partially address these issues. We continue our previous philosophy for the Amber "SB" (Stony Brook) force fields, assuming that physics-based force field development can provide excellent models with good transferability beyond their training data. Different approaches also have merit, such as in CHARMM, where physics-based training is followed by iterative rounds of empirical

adjustments that improve match to experimental data[12, 49]. The recent a99SB-disp model[29] derives from our ff99SB, followed by extensive empirical refitting of torsion parameters, nonbonded pair interactions, atomic partial charges and water dispersion energetics in order to improve agreement with experiments. Likewise, the recent ForceBalance approach is a promising method to automate iterative improvement through iterative cycles of fitting and comparison to experiment[50]. These adjustments can significantly enhance agreement with experiment, but the complex mapping of experimental observables to individual force field terms can also lead to the introduction of fortuitous (and non-transferable) cancellation of error between the various force field components. We attempt here to overcome the ff14SB weaknesses discussed above by a more self-consistent reconsideration of the physics-based training of protein backbone energetics, developing improved backbone parameters based on fitting to a wider variety of high-level QM, and eliminating a series of inconsistencies in past fitting that are likely to have negatively impacted the resulting models.

The first departure from ff14SB relates to the training data and the function for backbone energies. Here we fit the entire φ/ψ 2D QM energy surface (generated using 2D φ/ψ conformational scans), rather than just local minima as in ff99SB/ff14SB. This will eliminate the problem of unconstrained energies outside the energy minima. We used a 2D function to fit this surface, rather than independent 1D cosine terms for φ and ψ. This explicitly accounts for coupling between these correction terms. As shown in **Results**, the correction profile needed to match the ff14SB MM to QM for the ψ rotation differs depending on the value of φ. In other words, in ff99SB/ff14SB, it is not possible to use a 1D correction profile to accurately reproduce QM energy profiles for ψ at all values of φ and vice versa. This 2D "CMAP" approach was pioneered in the CHARMM force field[51], and extended here. The CMAP approach also was used

for backbone fitting in RSFF2+CMAP[44b], but in that case the free energy surface derived from PDB statistics was used as the fitting target, rather than QM data as we use here. Previously, the "CMAP" approach was employed by other Amber force fields as well. In ff99IDPs/ff14IDPs[28b, 45], the 2D energy profile was fit against statistical data from a PDB coil library. In ff12SB-cMAP[38], the CMAPs adjusted only the relative depths of minima (such as α basin vs. β basin) adjusted using a CMAP by fitting to helical propensities and β strand population in MD.

The second difference from ff14SB is that we address the polarization inconsistency during dihedral parameter fitting. While fitting the entire gas phase surface using CMAPs would ensure sampling of energies for regions populated in solution, a significant problem arises during dihedral fitting when comparing *in vacuo* energies between QM and MM. The MM fixed partial charges in most non-polarizable Amber models are traditionally fit to HF-level QM, which results in partial charges larger than expected in the gas phase, intending to mimic the higher dipoles induced in aqueous solution and avoid the need to explicitly include polarization in the FF calculation[48]. However, using these "pre-polarized" charges during dihedral fitting to reproduced higher level QM *gas phase* energies introduces error, and enforcing a match results in dihedral parameters that (at least partially) cancel out the effect of charge polarization. The ff03 Amber model addressed this by fitting new charges to QM calculations in low-dielectric organic solvent[52], but the subsequent protocol for backbone dihedral fitting (also in organic solvent) resulted in erroneous double-counting of solvation effects[53]. The recent "ipq" force fields[21, 23] addressed polarization inconsistency by using two independent charge sets, one for MD, fit to QM calculations that included a specific explicit water model[54] that was used in MD simulations, while a second set of gas-phase partial charges was used during fitting dihedrals corrections to gas-phase QM rotational energy profiles. Our approach differs; we train backbone

dihedrals using the same pre-polarized MM charges as used in MD, but using continuum aqueous solvation rather than gas-phase energies, and with reference QM data also in aqueous implicit solvent to resolve the gas/aqueous phase inconsistency (following precedent in RNA parameter fitting[53, 55]). An additional benefit of fitting against QM data in solution is that the target energy surfaces will reflect conformation-dependent changes in solvent-induced polarization, which are then incorporated into the CMAP even though the MM model uses fixed partial atomic charges.[56] These effects are missing in other current models (such as the "ipq" models) since dihedral fitting is done in the gas phase[21, 23].

More accurate reproduction of the QM training surfaces and resolving polarization inconsistencies allow us to undertake **the third difference** from ff14SB, that of exploring amino-acid specific correction maps. Amber already used separate parameters for proline and glycine, and finer differentiation is a reasonable next step. We believe that optimizing amino-acid specific backbone parameters using uncoupled functions for φ and ψ (as done by other groups[23, 28a]) is unlikely to result in significant improvement for ff14SB since these are not able to accurately reproduce the QM training data even for a single amino acid (see **Results**). For example, despite fitting sets of uncoupled cosine parameters for several groups of amino acids, simulations using the ff15ipq[23] force field show reduced accuracy for β-branched amino acids[23].

Alanine and valine (together with other β-branched isoleucine and threonine) are conformational outliers, justifying separate CMAP treatment. Alanine is very helical, whereas valine has a very flat φ distribution according to PDB φ/ψ distributions (**Figure 2**). Many residues exhibit conformational preferences between those of alanine and valine. Leucine is likely a better model for most amino acids (since all but Ala and Gly include a γ-carbon). We therefore used the CMAP fit to Leu for several other amino acids, including those with aromatic

rings (Phe, Trp, Tyr) and nonpolar but non- β-branched side chains (Met) and the three protonation states of His (His^+, His^δ, His^ε). Polar or charged side chains (Ser, Cys, Thr, Asp^-, Asp, Asn, Glu^-, Glu, Gln, Arg^+, Lys^+) all received individual CMAPs, Pro received its own CMAP and the β-branched Ile used the CMAP fit to the similar Val. Other force fields also fit different parameters for different amino acids. For example in Amber fb15[57], full scanning over φ/ψ and χ1/χ2 dihedrals were performed for each amino acid, then the 4D φ/ψ/χ1/χ2 grid was mapped onto 2D φ/ψ grid by searching for lowest energy side chain conformation at each φ/ψ. Then, uncoupled (1D) cosine functions were used for each dihedral φ, ψ, χ1 and χ2, with all phases and amplitudes fit simultaneously. Here, we fit 2D CMAPs to φ/ψ energy maps using a single rotamer for each amino acid, in order to avoid transferring errors in the χ energy profiles into the φ/ψ correction, as could happen if the φ/ψ grid points also vary in χ values.

Finally, we examine possible dependence of the backbone CMAP on side chain rotamer. In ff99SB and ff14SB backbone training (also CHARMM[12]), the coupling between backbone and rotamer was avoided by using the ff94 approach of Ala as a model for all other amino acids, thus ignoring any possible backbone-sidechain coupling correction. To account for rotamer dependency in RSFF2+CMAP[44b], the 2-dimensional ϕ/ψ CMAP was supplemented by the use of additional two-dimensional free energy surfaces including ϕ/χ1 and ψ/χ1. Here, we find that the 2D CMAPs that we fit to QM data in solution, in combination with the high-quality side chain energy profiles from ff14SB, result in a model that is reasonably transferable to side chain rotamers not included in the training data.

Extensive MD simulations (a total of ~5 milliseconds in explicit water) were performed to validate the performance of the ff19SB model. We show below that ff19SB, using amino-acid specific training against QM data with solvent polarization, reproduces the amino-acid specific

differences in crystallographic Ramachandran maps and experimental helical propensities much better than ff14SB or other older Amber models. We also show that the QM-based ff19SB is in reasonable agreement with experiments when combined with an accurate solvent model, while ff14SB performs poorly with the same solvent model and relies on cancellation of error with the less accurate TIP3P model in order to reproduce properties such as the helical content a Baldwin-type peptide. We conclude that an inherent underestimation of helicity is present in ff14SB, which is (inexactly) compensated by an increase in helical content driven that is likely driven by the reported[29-30, 58] TIP3P bias toward overly compact structures. The improvements in modeling helicity with ff19SB do not appear to result in less accurate performance on β systems. With ff19SB, the overall excellent performance of ff14SB and ff99SB in NMR order parameter reproduction is also generally maintained with even smaller RMSD values relative to experimental structures. Future work will examine the performance of ff19SB on IDP model systems.

Methods

Additional methods details are provided in Supporting Information.

Structure preparation & simulations

Unless noted otherwise, all crystal and NMR structures were downloaded from the PDB[59] at www.rcsb.org. Alternate structures including fully extended and fully helical used to initiate some simulations were built via the LEaP module of AmberTools in the Amber v16 software[4]. Helical and extended conformations are defined as (φ, ψ) = (-60°, -45°) and (φ, ψ) = (-180°, -180°) In explicit solvent MD simulations, TIP3P[60], OPC[32] SPC/E_b[24] and fb3[61] solvent models were used to solvate systems as noted. A truncated octahedron periodic box was used for all

simulations. Implicit solvent MD simulations with GBneck2 parameter set[62] of the GBneck solvent model[63] and ff14SB[10] were performed to generate additional initial structures. ff14SB[10], ff15ipq[23], fb15[57] and ff19SB were used for explicit solvent MD simulations as noted. System-specific details are discussed below with additional details in **Table S1**.

Dipeptides

Acetyl and N-methyl capped dipeptides of the natural amino-acids (Ace-X-Nme) were used for force field training and testing. In training, 16 amino acids (including two protonation states of Asp and Glu, but excluding Ile, Trp, Tyr, Phe, Met and His) were fully scanned in backbone dihedral space using implicit solvation (see **Structure preparation & simulations** and **Geometry scanning**). In testing, helical and extended conformations for all natural amino acids (including two protonation states each for Glu and Asp side chains, and three protonation states for His side chain) were used as initial structures in 800 ns MD simulations. The number of explicit water molecules was equalized across all dipeptide systems and solvent models (**Table S1**).This was achieved by adjusting the value of buffer distance until desired number of water molecules was obtained. Four combinations including ff14SB[10]+TIP3P[60], ff14SB[10]+OPC[32], ff19SB+TIP3P[60] and ff19SB+OPC[32] were tested for dipeptides.

Ala_5

Ala_5 with a free N- and protonated C-terminus was used in simulation, corresponding to pH=2 used in the NMR studies[43] (see **Parameter derivation for protonated C-terminal Ala**). Both helical and extended conformations were used as initial structures for 800 ns MD simulations. The number of water molecules was equalized across all runs (**Table S1**). Four combinations including ff14SB[10]+TIP3P[60], ff14SB[10]+OPC[32], ff19SB+TIP3P[60] and ff19SB+OPC[32] were tested for Ala_5.

$A_4\underline{X}A_4$ and $A_9\underline{X}A_9$ peptides

Acetyl and NH_2 capped polypeptides (matching pH=7 in NMR[39]) of the 20 natural amino-acids (A4**X**A4: Ace-A_4**X**A_4-NH_2 where **X** denotes the amino acid tested) were used to test amino-acid specific helical propensities. Two independent runs of 800 ns each starting from the helical and extended conformations were initially performed with ff14SB[10]+GBneck2[62], and cluster analysis (see **Cluster analysis**) was carried out on the combined trajectory. Cluster centroids from the top four clusters, together with helical and extended conformations were then selected as initial structures for MD simulations in explicit solvent (~4000 water molecules for both OPC and TIP3P runs, see **Table S1**). Each of these six initial structures seeded two independent runs each with different initial velocity assignment (using ig=-1 in Amber). Therefore, a total of 12 initial states were simulated for 3.2 μs each, in each explicit solvent, for each one of the 21 or 12 peptide sequences, for a total of 4146 μs for all force field + solvent model combinations (see **Table S1**). Helical propensities were calculated using six FF+water combinations including ff14SB[10]+TIP3P[60], ff14SB[10]+OPC[32], ff19SB+TIP3P[60], ff19SB+OPC[32], ff15ipq[23]+SPC/E_b[24] and fb15[57]+fb3[61].

Acetyl and NH2 capped polypeptides of the 20 natural amino-acids in a longer peptide (A9**X**A9: Ace-A_9**X**A_9-NH2 where **X** denotes the amino acid tested) were used to test the sensitivity of the helical propensities to chain length. Two independent runs, starting from helical and extended conformations, were initially performed for 800 ns with ff14SB+GBneck2, and cluster analysis (see **Cluster analysis**) was carried out on the combined trajectory. Cluster centroids from the top four clusters were then selected as initial structures for additional MD simulations in GBneck2. Each of these six initial structures seeded two independent runs with different initial velocity assignment (using ig=-1 in Amber). Therefore, a total number of 12

initial states were simulated in ff14SB[10]+GBneck2[62] for each one of the 20 Ace-A_9**X**A_9-NH2 systems, and each simulation was 2 μs long, for a total of 480 μs. These A9**X**A9 results were compared to data from A4**X**A4 (also in ff14SB[10]+GBneck2[62]) by extending the 800ns simulations described above to 2 μs.

K19 helical peptide

Consistent with our previous work[10, 64], the sequence of Ace-GGG(KAAAA)$_3$K-NH2 was chosen to validate parameter quality in folding helices. Since it was unfeasible to run long simulations starting from fully extended conformations that require very large numbers of water molecules to solvate, a fully extended conformation was not selected for explicit solvent simulations. For instance, 12000 water molecules would be needed to solvate a fully extended conformation of K19 with 8 Å buffer. Instead, several semi-extended initial conformations were generated. Two independent runs starting from helical and extended conformations were run for 800 ns with ff14SB[10]+GBneck2[62], and clustering analysis (see **Cluster analysis**) was performed on the combined trajectory. The cluster centroids (**Figure S1**) from the top 1st and 2nd were disregarded because both were partially helical with 2.7 Å and 4.4 Å RMSD (backbone C, N, CA atoms) referenced to a fully helical conformation. Therefore, the centroids from top 3rd (c2), 4th (3), 5th (c4) and 6th (c5) clusters were selected as semi-extended. Both semi-extended and helical conformations were immersed in explicit water. The number of water molecules was equalized across all runs (**Table S1**). Each initial structure was used for two independent runs with random initial velocity assignment (ig=-1 in Amber). Therefore, a total of 10 initial states were simulated with each force field + explicit solvent combination, and each simulation was 3.2 μs. Three combinations including ff14SB[10]+TIP3P[60], ff14SB[10]+OPC[32], and ff19SB+OPC[32] were tested for K19, for a total of 96 μs simulation.

CLN025 hairpin

CLN025 (PDBID: 2RVD[65], $^{+}H_3N$-YDPETGTWY-COO^{-}) is an engineered fast-folding hairpin that is a thermally optimized variant of Chignolin[66]. The native conformation was chosen as the 5th conformation in the NMR ensemble[65] since that conformation was closest to the average of the NMR ensemble. A fully extended conformation of the same sequence was also used, and four independent runs (ig=-1 in Amber) were performed with an explicit solvent for both native and extended conformations. Each simulation was 7.2 μs long and the number of water molecules was equalized across all runs (**Table S1**). Three combinations including ff14SB[10]+TIP3P[60], ff14SB[10]+OPC[32], and ff19SB+OPC[32] were tested for CLN025 for a total of 172.8 μs. A cutoff of 1.5 Å RMSD was chosen to delineate native from non-native structures because the highest population peak at low RMSD across all force field + solvent models ends near 1.5 Å (**Figure S2**).

Folded proteins

Three folded proteins were simulated for comparison to NMR-based backbone dynamics measurements. First was the third Igg-binding domain of protein G (GB3). The native structure was defined from a liquid crystal NMR structure (PDBID: 1P7E[67]). Second was Ubiquitin (Ubq), with the native structure defined from a crystal structure (PDBID: 1UBQ[68]). Third was hen egg white Lysozyme (HEWL), with the native structure defined from a crystal structure (PDBID: 6LYT[69]). Four independent runs with random initial velocity assignment (ig=-1 in Amber) were performed for each system in explicit solvent. Each simulation was 200 ns long and the number of water molecules was equalized across runs for each system (**Table S1**). These folded proteins were tested using three combinations including ff14SB[10]+TIP3P[60], ff14SB[10]+OPC[32], and ff19SB+OPC[32].

Geometry scanning

Backbone geometry scans were performed to generate structures for parameter training. All scans were carried out via the LEaP module of AmberTools in Amber v16 software[4, 70]. All 16 dipeptides (see ***Dipeptide*s**) were 2D scanned on φ and ψ dihedrals over ranges of -180° to 165° with an interval of 15°. For glycine dipeptide, a finer grid scanning with an interval of 5° was performed in the βregion (-180° to -125° and 120° to 175° on φ and ψ dihedrals) resulting in an additional 12*12 finer grid. We observed that the QM energy surface near βis highly sensitive to the selected grid point and using 15° interval might unintentionally miss the energy at the true minimum. For proline dipeptide, structures were limited to -180° to 120° on φ in order to exclude structures with excessive ring strain. For dipeptides containing one or more heavy atom χ dihedrals (Val, Leu, Ash, Asp^-, Asn, Glh, Glu^-, Gln, Lys^+, Arg^+, but excepting Ser, Cys and Thr, see below; Ash and Glh are neutral Asp and Glu, respectively), χ dihedral values were initialized to the most populated rotamer for that amino acid, according to Lovell's rotamer library[42].

Molecular mechanics (MM) optimization and energy calculations

For Cys and Met, Lennard-Jones (LJ) parameters were taken from GAFF2 for sulfur and hydrogen (in -SH and –S-) for CMAP fitting, and also incorporated into ff19SB. This was done to keep consistent with the most recent LJ parametrization on these atoms performed by Wang et al[71].

Unless otherwise noted, use of the term "GBSA" in this paper denotes the combination of GBneck2 (igb=8 in Amber) and SASA (gbsa=1 in Amber).

Dipeptide structures were minimized with restraints after geometry scanning. MM optimization and energy calculations were performed with Amber v16[4, 70] using ff14SB[10] and

GBneck2[62] implicit solvent model with the mbondi3 radii set[62] for polar solvation and SASA-based nonpolar solvation[72]. The default 0.005 *kcal mol^{-1} Å^{-1}* surface tension was adopted.

Dipeptides taken from geometry scanning were minimized using ff14SB[10] and GBSA including restraints on φ and ψ values with harmonic force constant of 1000 *kcal mol^{-1} rad^{-2}*. All χ dihedrals were relaxed during minimization without restraints, except Ser, Cys and Thr, for which the χ_2 dihedral (defined as CA-CB-OG-HG for Ser, CA-CB-SG-HG for Cys and CA-CB-OG1-HG1 for Thr) was restrained (10 *kcal mol^{-1} rad^{-2}*) to 165° to prevent the hydroxyl or sulfhydryl group from approaching too closely to the backbone amides during minimization. As we noted for ff14SB[10], this was done to avoid incorporating into the backbone dihedral parameters any difference between the quantum mechanical (QM) and MM models in the short-range potential between side chain and backbone.

We adopted the strategy of initializing all structures on the grid at the same rotamer conformation, then minimizing with backbone restraints to relax the rotamer to a local minimum. The rationale for using a single initial rotamer for the entire φ/ψ grid scan is to reduce the likelihood of transferring any errors in the ff14SB side chain rotamer energy profiles to the CMAP (which can occur if neighboring grid points also differ significantly in χ dihedral values). The same relaxed rotamer was used in the QM calculations (discussed below).

Structures were minimized for a maximum of 10,000 cycles in ff14SB+GBSA with no cutoff on non-bonded interactions. Steepest descent was employed for the first 10 cycles in the minimization and conjugate gradient for the following cycles. The convergence criterion for energy gradient is when the root-mean-square of the Cartesian elements of the gradient is less than 10^{-4} *kcal mol^{-1} Å^{-1}*. Single point energies were calculated for the MM-optimized structures using ff14SB00+GBSA. ff14SB00 is defined as the original ff14SB[10] force field with the

amplitudes of dihedrals sharing the same central two atoms with φ and ψ (C-N-CA-C, C-N-CA-CB, N-CA-C-N, CB-CA-C-N, HA-CA-C=O) set to zero (see the content of the Amber frcmod-format data in **Table S2**).

CMAP fitting groups

A total of 16 CMAPs were fit and then applied to the 20 natural amino acids with several having alternate protonation states (**Table S3**). CMAPs are only applied to central residues but not to uncapped terminal residues. Ala, Gly, Pro were fit separately because the allowable regions in Ramachandran plot according to PDB are notably different from each other.[73] Ser, Cys and Thr were fit separately from others because of the proximity of the polar group to the backbone, and from each other because the polarity of their side chains is different (Ser vs. Cys) or the side chain β-branching structure is different (Ser vs. Thr). Val CMAP was fit and applied to both Val and Ile since Val and Ile are the only two amino acids having β-branched non-polar side chain. Arg^+, Lys^+, Asp^-, Ash, Asn, Glu^-, Glh and Gln were fit separately because the charge state is different (Arg^+ and Lys^+ vs. Asp^- and Glu^-), the polarity of side chain is different (Arg^+ vs. Lys^+, Asp^- vs. Ash vs. Asn, Glu^- vs. Glh vs. Gln), or the length of side chain is different (Asp^- vs. Glu^-, Ash vs. Glh, Asn vs. Gln). Leu CMAP was fit and applied to long non-polar and non-charged side chains including amino acids with aromatic rings (Phe, Trp and Tyr), Met, Cys in disulfide bonds (Cyx) and Cys interacting with metal (Cym). Leu CMAP was also applied to the three protonation states of His (His^+, $His^ε$, $His^δ$).

CMAP fitting

A CMAP is defined by a 24*24 grid that is evenly spaced (15°) in φ/ψ dihedral space, the same spacing as used in C22/CMAP[40a], C36[13b], C36m[40b] and RSFF[28a, 44] force fields. At each grid point, the energy $U_{cmap}(\varphi, \psi)$ corresponds to the following:

$$U_{cmap}(\varphi,\psi) = E_{QM}^{gas} + E_{QM}^{polarization} + E_{QM}^{solvation} - (E_{MM}^{ff14SB00} + E_{MM}^{solvation}) \quad (1),$$

where E_{QM}^{gas} represents gas-phase QM energy, $E_{QM}^{polarization}$ represents the contribution from solute-solvent polarization from QM solvation and $E_{QM}^{solvation}$ represents the remaining specific solvation effects in QM. $E_{MM}^{ff14SB00}$ represents MM energy calculated in ff14SB00 (**Table S2**) using pre-polarized charges, and $E_{MM}^{solvation}$ represents MM solvation energy calculated in GBSA. In practice, E_{QM}^{gas}, $E_{QM}^{polarization}$ and $E_{QM}^{solvation}$ cannot be separated since the solute electron density is evaluated self-consistently with the solvent polarization represented in a reaction field. In Amber, the CMAP grid is represented using a bicubic spline function

$$U_{cmap}(\varphi,\psi) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}\varphi^i\psi^j \quad (2),$$

where φ and ψ are dihedral values in radians, and a_{ij} are coefficients that are solved from a set of linear equations derived from values at the 4 corners of the grid cell[74]. This function is used to interpolate MM energy at any arbitrary φ/ψ dihedral.

The resulting CMAP forces are calculated by the chain rule and added to the total forces[75]. The CMAPs are intended to be used as direct replacement for the cosine-based φ/ψ dihedral terms in ff14SB and should not be combined with those cosine terms. The CMAP code was originally implemented in Amber with the support of CHAmber module[75]. Further details are provided in Supporting Information.

QM energies in solution

To calculate $E_{QM}^{gas} + E_{QM}^{polarization} + E_{QM}^{solvation}$ (**Equation 1**), we used the SMD solvent model that includes both polar and nonpolar solvation components[76]. The polar component uses the integral-equation-formalism polarizable continuum model (IEFPCM)[77], where the solute cavity is defined through superposition of atom-centered spheres with reparametrized "intrinsic radii".

The non-polar component is a product of the solvent-accessible surface area (SASA) and the surface tension, which is a function of several element-specific parameters. These empirical parameters for effective radii and surface tension were iteratively optimized to reproduce 2346 solvation free energies of both neutral solutes and ions[76]. In the original work[76], the authors concluded that among various QM theories used in their parameter fitting, the DFT method M05-2X[78] yielded the best performance. Taking these results into consideration, particularly performance for amides, we selected the hybrid functional M05-2X with basis set 6-311G** together with SMD to compute the total solvation energy in QM. In the original paper[76], 6-31+G** was shown to have smaller mean unsigned error in aqueous solvation free energy for all tested molecules, including four amides, compared to other basis sets such as MIDI!6D, 6-31G* and cc-pVTZ. The diffuse functions in 6-31+G**, however, cause convergence issues in some of our calculations where the geometries are far from equilibrium. Instead, we use the comparable 6-311G** basis set. We also tested whether M05-2X/6-311G** is accurate in calculating conformational energies of Ala dipeptide in gas-phase (see **SI: Comparing various methods on QM energy calculations in gas phase**). The calculated average relative energy error (see **Methods: Average relative energy error (REE) calculation**) against MP2/cc-pVQZ for nine conformations of Ala dipeptide is ~0.35 kcal/mol, very close to MP2/cc-pVTZ level of accuracy (average REE = 0.2 kcal/mol). Based on our results, M05-2X/6-311G** is reasonably accurate relative to MP2/cc-pVQZ at reproducing relative energy of Ala dipeptides in gas phase (**Figure S3**), and errors are likely comparable to those arising from other sources such as the spacing of the grid scan and fundamental inaccuracies in the MM treatment.

QM optimization and energy calculations

QM calculations were performed with Gaussian 09[79]. Geometry optimizations and single point energy calculations were performed on the 16 dipeptides at the M05-2X/6-311G**/SMD level of theory[78]. Grimme's dispersion correction with the original D3 damping function[80] was used to correct for long-range dispersion. The solvation environment was represented as a self-consistent reaction field, with exterior dielectric set to default 78.3553, using SMD[76] with consideration of both polar and nonpolar solvation energy components.

Very tight optimization convergence criterion was used to generate data for fitting. To maintain the structure on the φ/ψ grid, one of the dihedrals sharing the same central two atoms with φ, and one dihedral sharing the central two atoms with ψ were constrained to the values from the structures taken from the last step of MM optimization. In order to avoid inclusion of errors in the χ energy profiles into the QM-MM energy difference used for CMAP fitting, we also restrained one of the dihedrals for each χ dihedral to the value from the last step of MM optimization (see **Molecular mechanics (MM) optimization and energy calculations**) (details on restrained dihedrals provided in **Table S4**).

For glycine dipeptide, QM optimization and energy calculations were done on both 24*24 coarse grid with 15° interval (same as other amino acids) and 12*12 sub-grid with 5° interval (specific to glycine). In the region with fine grid data, the QM energies of coarse grid points were replaced with lowest energy in the surrounding fine grid points (within 10° from the coarse grid). This was only done for QM energy calculations on Gly. MM calculations on Gly were performed the same as other amino acids, and the 24x24 grids were used to obtain the CMAPs.

Average relative energy error (REE) calculation

Unless otherwise noted, average REE between two sets of energies were calculated as following:

$$average\ REE = \frac{2}{N*(N-1)}\sum_{i}^{N-1}\sum_{j>i}^{N}\left|\left(E_i^a - E_j^a\right) - \left(E_i^b - E_j^b\right)\right| \quad (3),$$

where N is the number of conformations. E_i^a and E_j^a are energies calculated in method "a" (QM, MM, etc) of conformation i and j. E_i^b and E_j^b are energies calculated in method "b" (QM, MM, etc) of conformation i and j.

Parameter derivation for protonated C-terminal Ala

Following the original RESP method for peptide partial charge assignment[6b, 48b], new charges were fit using the RESP approach for an Ala residue with a protonated C-terminus. The model system for fitting was Ala with a protonated C-terminus and an acetylated N-terminus (using existing charges for this capping group). Helical and extended conformations were used for RESP fitting. The partial charges on all atoms except the –COOH group were restrained to the charges from ff94[48a] while the –COOH charges were refit via RESP. QM calculation was performed with Gaussian 09[79]. HF/6-31G* was used for geometry optimization. MK[81] population analysis was performed on the optimized geometry. Antechamber, espgen and residuegen as implemented in Amber v16[70] were used in RESP fitting.

The resulting atomic charges are listed in **Table S5**. The –COOH functional group in the protonated C-terminal Ala was assigned the same atom types as –COOH in side chains of Alh or Glh, thus sharing existing bonds, angles, dihedral and LJ parameters. When simulating a system with a protonated (uncapped) C-terminal Ala in ff19SB, standard ff14SB parameters were applied to the C-terminal residue without application of a CMAP due to lack of C-terminal amide.

MD simulations

The following methods were used for all MD simulations unless otherwise noted. Bonds to hydrogen atoms were constrained with the SHAKE algorithm[82] using a geometrical tolerance of

0.000001Å. The direct space non-bonded interaction cutoff was 10.0 Å for explicit solvent simulations and 9999.0 (no cutoff) for implicit solvent simulation. Long-range electrostatic interactions in explicit solvent were calculated via the particle mesh Ewald (PME) approach[83]. There were a total of 9 steps of equilibration in both implicit and explicit simulations (see **SI: MD equilibration**). For production runs, the time step was increased to 4 fs using the hydrogen-mass repartitioning method implemented as described previously[84], and explicit solvent simulations were changed to the NVT ensemble (ntb=1, ntp=0 in Amber).

Cluster analysis

Unless noted otherwise, cluster analysis was performed on the combined trajectories starting from helical and extended conformations. The hierarchical agglomerative (bottom-up) approach was used with average linkage (defined by RMSD of C, N and CA atoms) to generate a maximum of 10 clusters using default settings in Cpptraj[85]. Our goal was to identify alternate structures to initiate independent runs, thus we specified the number of clusters rather than a similarity metric for cluster members since structures other than the representative were not examined further. The representative structures extracted from these clusters were used as initial conformations in independent MD runs to check convergence.

RMSD calculations

Unless otherwise noted, all RMSD calculations were done on backbone C, N and CA atoms via Cpptraj[85]. In all cases, terminal residues and capping groups on termini were neglected.

Helical propensity

Following the Best et al.[13a] protocol of implementing Lifson-Roig model[86] for computing helical propensities, we explored the helical propensities of each amino acid in the context of the sequence Ace-A_4**X**A_4-NH_2 to compare to experimental data[39]. This model measures the

equilibrium properties of coil-to-helix transitions: three states are defined: coil, start/end of the helix, and within a helix. Their relative weights are 1, v_i *and* w_i, respectively. The start/end of the helix is defined when residue *i* is in the helical region but either of its two adjacent residues is not in the helical region. The residue within a helix is defined when residue *i* and its two adjacent residues are all in the helical region. Everything else is considered to be random coil within the model. A residue is considered helical if inside the α region using the basin definition in **Table S6**. The sensitivity to this definition was tested by calculating helical propensity with a wider range definition (**Table S6** and **Figure S4**).

Following Best et al.[13a], the partition function for the blocked peptide of length N (N=9) is defined as:

$$Z = (0\ 0\ 1)\prod_{i=1}^{N} M_i\ (0\ 0\ 1)^T, \text{ where } M_i = \begin{vmatrix} w_i & v_i & 0 \\ 0 & 0 & 1 \\ v_i & v_i & 1 \end{vmatrix} \quad (4),$$

The log-likelihood that residue *i* will be assigned a helical propensity parameter w_i is given by:

$$ln(L) = \sum_i N_{w,i} ln(w_i) + \sum_i N_{v,i} ln(v_i) - N_k ln(Z) \quad (5),$$

where v_i and w_i are the parameters for fitting, N_k is the total number of frames in the simulation, $N_{w,i}$ and $N_{v,i}$ are the total number of times in the simulation that residue *i* is within a helix and start/end of a helix, respectively. More specifically, $N_{w,i}$ is incremented if residue *i* is within a helix and $N_{v,i}$ is incremented if residue *i* is start/end of a helix. The subscript *i* indicates the amino acid (Ala, Val, Leu, etc). The model parameters (*v* and *w*), and their distributions, were optimized by following the genetic algorithm protocol of Perez et al.[38] to maximize the objective function, *ln(L),* which maximized the likeliness of residue *i* being assigned to specific *v* and *w*. Mutation and crossover moves were performed to change $ln(w_i)$ and $ln(v_i)$, with a rate of 0.3 and 0.7 respectively. A total of 1000 genetic optimization cycles were performed to yield

specific v and w for residue i. v_{ala} and w_{ala} were initially evaluated for all Ala in the capped $A_4\underline{\mathbf{A}}A_4$ peptide, then v_i and w_i for $\underline{\mathbf{X}}$ were evaluated in capped $A_4\underline{\mathbf{X}}A_4$ peptide with v and w parameters for Ala being fixed to the values of previously optimized v_{ala} and w_{ala}. The helical propensity data (from both NMR and MD) are provided in **Table S7** and **Table S8**.

Bootstrapping analysis on helical propensity

In order to quantify the uncertainty of the computed w, bootstrapping analysis was performed for each system. When the sample size is insufficient for straightforward statistical inference, bootstrapping provides a way to account for the distortions caused by a specific sample that may not be fully representative of the population. First, a combined trajectory from 12 independent runs of each $A_4\underline{\mathbf{X}}A_4$ (3.2 μs for each run) was used to fit the v and w for that $\underline{\mathbf{X}}$. Second, the combined trajectory was split into 10 segments with same length. Third, 50 times of resampling with replacement were done on the 10 sub-trajectories. This step generated 50 trajectories with the same length of the initially combined one (3.2 μs * 12) but with some segments repeated. 50x resampling has been suggested to lead to reasonable standard error estimates[87]. Lastly, we fit the v and w parameters with each of the 50 trajectories respectively and calculated the standard deviation of the 50 resulting w values for each amino acid. According to the distribution of the sampled w parameters (**Figure S6**), all amino acids have a high peak and a narrow range on w which suggests good quality sampling and precise estimates of helical propensity.

NMR scalar coupling calculations

Following Best et al.[15] and our previous work[88], scalar couplings were calculated from simulations using Karplus relations[89] and the "Orig parameters"[90] also adopted by Graf et al.[43].

To quantify the discrepancy between experimental scalar couplings and those calculated from simulation, χ^2 error was defined in Best et al.'s work[15] and also here as:

$$\chi^2 = N^{-1} \sum_{j=1}^{N} (\langle J_j \rangle_{sim} - J_{j,exp})^2 / \sigma_j^2 \quad (6)$$

where N is the total number of scalar coupling types, $\langle J_j \rangle_{sim}$ is the averaged scalar coupling from the simulation for scalar coupling type j. $J_{j,exp}$ is the NMR data for type j. σ_j is the estimated systematic error of the Karplus equation for type j adopted by both Best et al.'s work[15] and our previous work[88]. Precision of χ^2 is estimated as half the difference of χ^2 calculated from two simulations starting from either helical or extended conformation. For dipeptides, **Table S9** lists ${}^3J_{HNHA}$ data and an estimated systematic error of 0.91 was used in the χ^2 calculation[15, 88]. For Ala_5, **Table S10** lists all scalar coupling types and the corresponding systematic errors[10]. Since the NMR data[91] were measured at pH=4.9, side chains for Arg, Lys and His were modeled in protonated state. For Glu and Asp, both deprotonated and protonated states were simulated, and the error was reported as a weighted average value. Constant pH simulations were performed to obtain the appropriate ratio of protonated state versus deprotonated state respectively.

Constant pH simulation

Constant pH simulations of 800 ns with TIP3P[60] and OPC[32] explicit solvent were performed on the capped dipeptide forms for the titratable residues Glu and Asp. Initially, these titratable residues were assigned to be protonated, and the state change was attempted every 100 MD steps through Monte Carlo approach using a Generalized Born implicit solvent model (igb=2)[92] which was the model used to parameterize the reference compounds in constant pH simulation[93]. Following published protocol, the intrinsic Born radii of carboxylate oxygen atoms were shrunk in order to reduce artifacts arising from including all four alternate hydrogen atom positions in the GB descreening calculation.[93] 200 steps of explicit solvent relaxation dynamics (in which the

solute was held fixed) were performed before resuming simulation if any protonation states were changed[93]. The solvent pH value was set to 4.9 in agreement with the NMR experiment[91]. The rest of the input was retained from the standard MD protocol described above.

These constant pH simulations have limitations, such as using an older GB model[92] (igb=2) for reference compound energy, and neglect of updating dihedral parameters when protonation state switches[93]. Therefore, the constant pH simulations were only used to estimate the percentage of protonation states for titratable residues, and the sampled ensembles were not used directly for χ^2 analysis. The χ^2 analysis was performed on the combination of protonated and deprotonated trajectories in explicit water, weighted by the ratio of protonated state versus deprotonated state obtained from constant pH MD.

NMR order parameters

The ability of a force field to model local dynamics accurately in well-folded proteins in solution was examined by comparing to NMR experimental backbone NH S^2 order parameters for GB3[67], ubiquitin[68] and lysozyme[69]. We adopted the model-free approach originally proposed by Lipari and Szabo[94] and used iRED[95] as implemented in Cpptraj. iRED is based on a covariance matrix analysis of inter-nuclear vector orientations, represented by spherical harmonics, extracted from MD simulations. For this analysis, we averaged iRED results calculated for windows of length five times the tumbling correlation time (τ_c), which was suggested to best reproduce the model-free S^2 order parameters[96]. Thus, window sizes of 2 ns, 4 ns and 8 ns were used for GB3, ubiquitin and lysozyme respectively, in agreement with previous work[97]. The uncertainties in the computed S^2 were calculated by taking the standard deviation from four independent MD runs. The absolute difference (AD) is computed between theoretical and experimental S^2 for each residue.

Statistical analysis of PDB data

To compare the φ/ψ distributions from simulation against PDB data, we used Lovell's rotamer library[42, 98] of 7957 high-resolution, quality-filtered protein chains to generate the PDB-based φ/ψ distributions. Two filters were applied to select a portion of the original 7957 structures. Firstly, only residues in coil and turn as defined by DSSP[99] were selected. Secondly, these residues were eliminated if any of the backbone heavy atoms had B factors larger than 30. Biopython[100] was used to apply the two filters against 7957 PDB files (PDB IDs and residue numbers of the filtered structures are provided in **SI file pdbid_residue.csv**. The script used to generate the filtered results are provided in **SI file filtering_extract_phipsi.sh**).

Results and Discussion

Backbone rotational energies in ff19SB compared to ff14SB

Alanine and Glycine energetics. Backbone φ/ψ rotational energy profiles were analyzed for QM, ff19SB, ff14SB and CMAP (derived by subtracting ff14SB00 from QM energies on the 2D grid, see **Methods**). Ala and Gly are discussed first because they are the simplest with no significant side chain degrees of freedom. We performed 2D backbone rotation scans for the capped Ala and Gly dipeptides, followed by restrained minimization and energy evaluation with implicit solvent for QM and MM. The CMAPs were derived by subtracting MM from QM energies on the 2D grid. The ff19SB energies were obtained by adding the CMAP-based bicubic function to ff14SB00 (see **Methods: Molecular mechanics (MM) optimization and energy calculations** and **CMAP fitting**). As shown in **Figure 3**&**4**, the ff19SB energy profiles are nearly identical to the QM reference data, which was anticipated based on the training method. However, significant differences between ff14SB and QM are apparent. In ff14SB, the overall

energy profiles are highly symmetric with little φ/ψ coupling, likely due to the lack of coupling between the ff14SB dihedral correction parameters. This coupling may arise from polarization changes as the amide dipoles become aligned in the helical conformation. The shape and location (the bin having lowest energy in the basin defined in **Table S11**) of the α basins from QM are poorly reproduced by ff14SB for both Ala and Gly. Importantly, the diagonal shape of the left- and right-handed α helical basins as observed in QM and ff19SB is poorly reproduced in ff14SB, which instead samples too deeply into negative φ for ψ < 0. In addition, for Ala, the C_7^{eq} local minimum between ppII and α_R in QM (**Figure 3**) is absent in ff14SB, but reproduced with ff19SB. For Gly, the QM energy barrier at φ = -120 / ψ = -60 is more accurate with ff19SB (**Figure 4**).

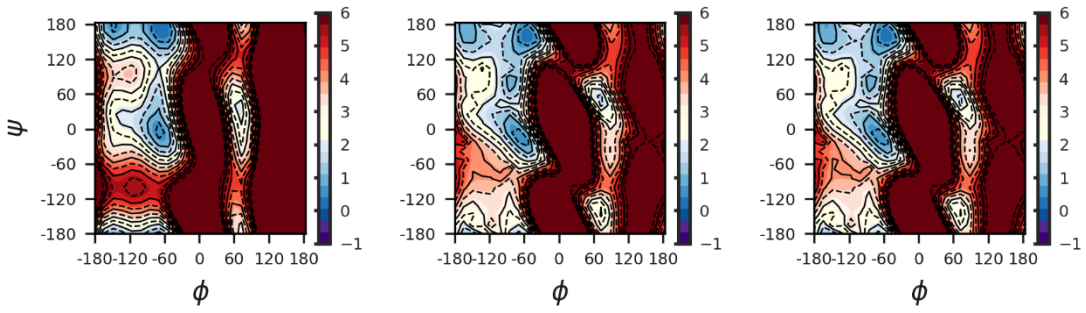

Figure 3. Ala dipeptide Ramachandran potential energy surfaces (kcal/mol) calculated in (left) ff14SB+GBSA, (middle) QM+SMD and (right) ff19SB+GBSA. All energies were zeroed relative to the lowest energy in the ppII region (defined in **Table S6**). The values above the color bar range are depicted in dark red. Solid, labeled contours indicate integer energy values in kcal/mol, whereas dashed contours indicate half-integer energies. The bicubic spline interpolation implemented in Python was used to calculate values between grid points.

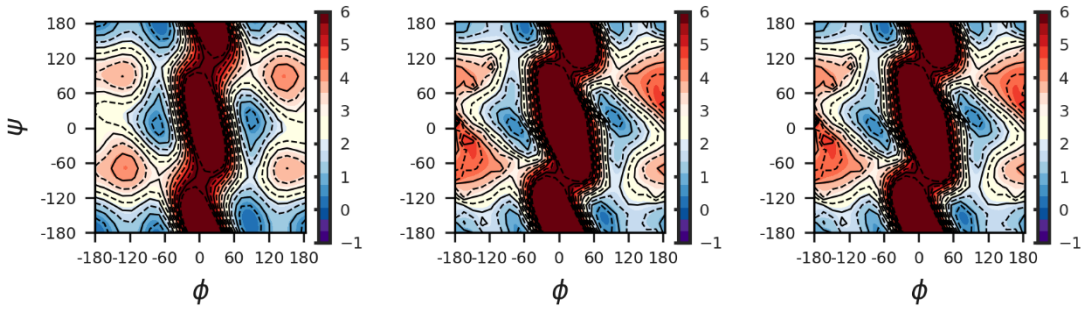

Figure 4. Gly dipeptide Ramachandran potential energy surfaces (kcal/mol) calculated in (left) ff14SB+GBSA, (middle) QM+SMD and (right) ff19SB+GBSA. All energies were zeroed relative to the lowest energy at ppII region (defined in **Table S6**). The values above the color bar range are depicted in dark red. Solid, labeled contours indicate integer energy values in kcal/mol, whereas dashed contours indicate half-integer energies. The bicubic spline interpolation implemented in Python was used to calculate values between grid points.

Overall, the deviation of ff14SB from QM for Ala and Gly is notable despite the use of QM data for multiple conformations of Ala_3 and Gly_3 during training of ff14SB/ff99SB backbone parameters. This relative weakness in ff99SB/ff14SB is likely a result of the use of only gas-phase energy minima for training (thus lacking the compulsion to reproduce the entire basin shape, or even the locations of aqueous-phase minima), along with dihedral correction terms that lack φ/ψ coupling, resulting in an overly symmetric energy map. Use of the CMAP approach for ff19SB results in improved reproduction of the overall energy surfaces for both amino acids.

We tested the impact of using QM in gas-phase as the target data. We fit an Ala dipeptide CMAP (same protocol as in **CMAP fitting**) against the grid surface of gas-phase QM energy instead of aqueous QM (**Figure S7A**), and ff14SB00 was used for gas-phase MM energy. The resulting energy surface (ff_gas) applied in solution (**Figure S7C**) has an unusual shape of the α_R

basin (extending much farther into ϕ<-120°), and the α_L energy basin is unexpectedly deep as compared to the results obtained using the aqueous QM surfaces. We conclude that fitting CMAPs using solution QM & MM calculations is important for good results here.

Amino acids with multiple side chain rotamers. The 2D CMAP training provides a "perfect" fit against the 2D reference QM data for Ala and Gly since no other significant rotational degrees of freedom are present. However, all other amino acids have longer side chains with additional degrees of freedom, and the situation becomes more complex since the energies (and their errors) depend on rotational degrees of freedom not sampled explicitly during CMAP training. While 3D fitting might accommodate some amino acids such as Val or Ser, this rapidly becomes intractable. We first compare alanine and valine using the valine rotamer used in training, then evaluate the transferability of the Val CMAP to alternate Val rotamers.

Our strategy to improving rotamer dependence extends the approach to improving transferability in the side chain parameters we used when developing ff14SB, where we assumed that the largest contribution to poor transferability of dihedral parameters arises from including structures in the training set that expose inaccuracies in the MM short-range nonbonded model that depend on degrees of freedom outside those being trained. If correction to these errors were to be incorporated into the backbone parameter for that φ/ψ grid point, it would be applied erroneously for conformations sampling the same φ/ψ values but with different rotamers that lack these inaccurate interaction energies. Therefore, rotamer dependency was addressed here by initializing all structures on each CMAP training φ/ψ grid at the same rotamer conformation, then locally relaxing the side chain conformations to relieve any backbone:sidechain steric clashes that were likely to be inaccurately modeled in MM. If corrections for training set structures with inaccurate backbone:rotamer MM energies were to be incorporated into the

backbone parameter for that φ/ψ grid point, the CMAP would have poor transferability to structures with the same φ/ψ values but with alternate rotamers that lack these inaccurately modeled interactions (see **Methods: Molecular mechanics (MM) optimization and energy calculations**).

Comparison of Alanine and Valine Energy Surfaces. For Val, we selected the *trans* rotamer for CMAP training (**Figure 5 first row**). As shown in **Figure 3** and **Figure 5B**, the QM profiles are qualitatively different between Ala and Val. Val prefers a flatter β/ppII transition region with a U-shape, while Ala has a higher barrier, a stronger preference of ppII over β, and a lower transition barrier between αR and ppII. The C_7^{eq} local minimum between ppII and αR observed in Ala is absent in Val. In addition, the elongated diagonal shape of the αR and αL basins in Ala (indicating strong φ/ψ coupling) is quite different from the narrow circular minimum in Val. The energy minimum at φ = 60 and ψ -150 in Ala is shifted upwards at φ = 70 and ψ = -60 in Val. Importantly, these differences in the Ala/Val QM surfaces are reproduced poorly in ff14SB where the Ala and Val surfaces are generally too similar; both Ala and Val prefer ppII over β and have similar symmetric α_R/α_L basins (**Figure 3(left)** vs. **Figure 5A**).

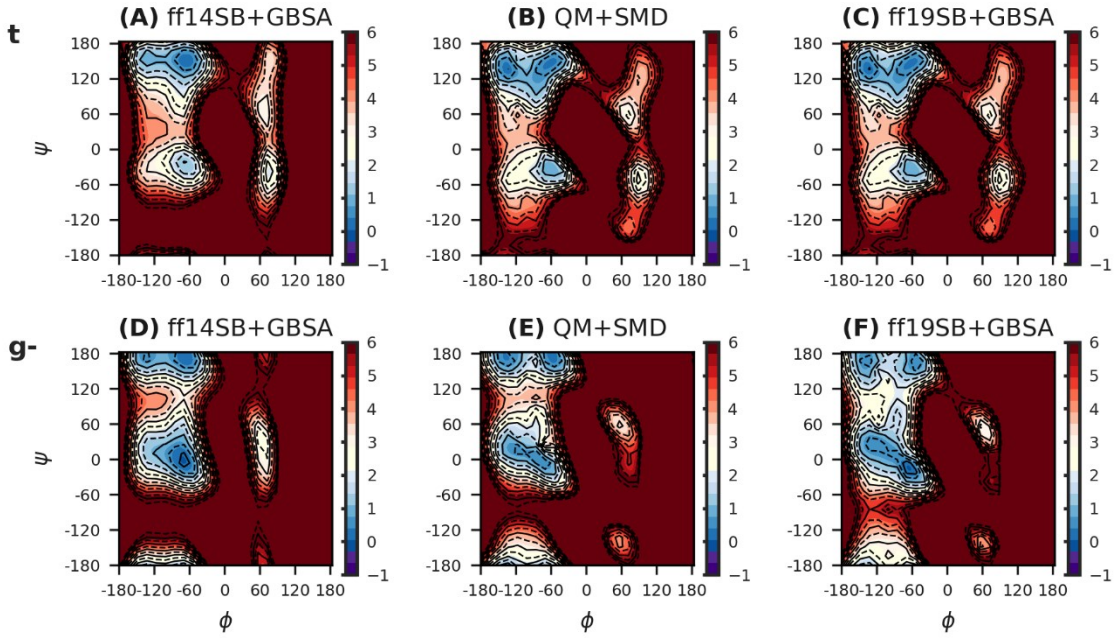

Figure 5. Val dipeptide Ramachandran potential energy surfaces using the *trans* (t) rotamer, calculated in (A) ff14SB+GBSA, (B) QM+SMD and (C) ff19SB+GBSA, and using the *gauche(-)* (g-) rotamer, calculated in (D) ff14SB+GBSA, (E) QM+SMD and (F) ff19SB+GBSA. The *trans* rotamer was used for ff19SB training. All energies were zeroed relative to the lowest energy at ppII region (**Table S6**). The values beyond the color bar range are depicted in dark red. Solid, labeled contours indicate integer energy values in kcal/mol and dashed contours indicate half-integer energies. The bicubic spline interpolation implemented in Python was used to calculate values between grid points.

Transferability of ff19SB backbone parameters to different side chain rotamers

We tested the ability of our approach to provide reasonable transferability of CMAPs between alternate rotamers using valine, for which the side chain rotamer is known to significantly influence backbone populations[17, 42, 101]. We switched the Val rotamer from *trans* to *gauche(-)*,

calculating QM and MM φ/ψ energies for *gauche(-)* conformations, but keeping the Ala-based ff14SB and *trans*-based Val ff19SB MM parameters (**Figure 5, bottom row**). Even though ff19SB was fit using the *trans* rotamer, it reasonably reproduces the changes in the Val QM data from *trans* to *gauche(-)*. For example, moving from *trans* to *gauche(-),* the α basins become more diagonal, α_L extends farther into the upper left quadrant, the barrier between ppII and β increases, and the minimum at (90, -60) disappears. As seen with the Ala/Val comparison, ff14SB poorly reproduces each of these changes, and the overall energy profiles are generally much too similar between the two rotamers, inconsistent with the QM results. Even though the α basin is stabilized more and becomes wider from *trans* to *gauche(-)* for ff14SB, the energy profiles are still highly symmetric in both rotamers and the notable difference in the shape of α basins reflected by QM and ff19SB is poorly reproduced in ff14SB, along with a too-flat barrier between ppII and β. Furthermore, rather than the disappearance of the (90, -60) minimum as seen in QM and ff19SB, the two minima with positive φ values merge into a single minimum in the wrong location with ff14SB. Thus even though ff19SB was trained using a single rotamer for Val, it does a better job than ff14SB at reproducing the rotamer-dependent backbone profiles from the QM calculations. The results also demonstrate that the high quality match between QM and ff19SB is not simply the result of empirical fitting to an energy map with a single rotamer, but that the accurate reproduction of the QM profiles is maintained even when the map is qualitatively different for an alternate rotamer. To quantify the changes, we calculated average REE (see **Methods: Average relative energy error (REE) calculation**) between QM and MM for *trans* and *gauche(-)* as a function of QM energy range above the minimum (**Figure S8**). For structures having QM energy within 7 kcal/mol above the minimum, the average REE for the training rotamer *trans* are 1.78 kcal/mol and 0.03 kcal/mol for ff14SB and ff19SB respectively.

The average REE for the test rotamer *gauche(-)* are 1.39 kcal/mol and 0.89 kcal/mol for ff14SB and ff19SB. Reasonable transferability is observed for other amino acids as well; examples include Ser and Glu. For Ser (**Figure S9**), ff19SB was trained against *gauche(+)*, but is able to reproduce reasonable QM surfaces for both *gauche(+)* and *gauche(-)*, such as the diagonal shape of α_R and α_L basin for both rotamers and the local minimum between ppII and α_R for *gauche(-)*. For structures having QM energy within 7 kcal/mol above the minimum, the average REE for *gauche(+)* are 1.80 kcal/mol and 0.06 kcal/mol for ff14SB and ff19SB. The average REE for *gauche(-)* are 1.98 kcal/mol and 1.01 kcal/mol for ff14SB and ff19SB. For Glu (**Figure S10**), ff19SB was trained against rotamer mt-10 (using naming conventions from literature[42]) (*gauche(-)* for $\chi1$, *trans* for $\chi2$ and -10° for $\chi3$) and reproduces reasonably the QM surfaces for both mt-10 and tt0 (*trans* for $\chi1$, *trans* for $\chi2$ and 0° for $\chi3$). In contrast, ff14SB merges the two minima into one at $\phi = 60°$ for mt-10, and poorly reproduces the barrier height at $\phi = -120°$ and $\psi > 30°$ for tt-0[42]. For structures having QM energy within 7 kcal/mol above the minimum, the average REE for mt-10 are 2.05 kcal/mol and 0.08 kcal/mol for ff14SB and ff19SB. The average REE for tt10 are 1.82 kcal/mol and 0.72 kcal/mol for ff14SB and ff19SB.

The QM, ff14SB and ff19SB energy maps for all 16 amino acid dipeptides in the training set are shown in **Figure S11**.

Amino-acid specific Ramachandran sampling from PDB is reproduced better with ff19SB

As shown above, the CMAP procedure allows the MM 2D φ/ψ energy surfaces to quantitatively match the QM 2D training data. Furthermore, we showed that using CMAPs improves the ability of MM to reproduce changes in QM φ/ψ basin shapes and locations for different χ rotamers. An important question, though, is whether these QM-based training data for

dipeptide models in solution provide good reference states for longer peptides in solution, or larger proteins with more complex structures and interactions. In order to explore the relevance of the differences seen between the ff19SB and ff14SB energy maps for different amino acids, we sought out high-quality PDB data[42, 98] on each amino acid and compared them to dipeptide φ/ψ sampling in MD using ff19SB. As discussed above in the context of statistical potentials, such comparisons have significant flaws, largely arising from the imperfect assumption that the distribution of backbone conformations for an amino acid across different proteins in a crystal environment (at different and typically low temperatures) corresponds to the MD-sampled Boltzmann distribution for the unconstrained peptide in solution at room temperature. Here, we restrict the use of the PDB data to a comparison of qualitative differences between amino acids from the same data source, such as from PDB or MD simulations. We expect that comparison of general features such as simulation and crystallographic basin shapes could provide valuable feedback that is independent of the dipeptide QM training data. However, we avoid assessment of quantitative features such as basin energies, for the reasons discussed above.

Distributions from the high resolution crystal structures[73] ("PDB"), dipeptide MD in ff14SB+OPC and dipeptide MD in ff19SB+OPC are shown in **Figure 6** for Ala, Val and Leu (with all amino acids shown in **Figure S12**). The OPC solvent model was selected for this test since this model was developed by optimizing the charge distribution to match QM data and vdW parameters to reproduce water density. Neither ff14SB nor ff19SB parameters were empirically adjusted with OPC (ff14SB used TIP3P-based MD data in training).

Because the dipeptide is fully exposed to the solvent, the results are more sensitive to the protein force field than to the solvent model, consistent with our previous report that similar populations of basins of Ala dipeptide were obtained with various solvent models[102]. Here,

similar distributions for each amino acid are observed between ff14SB+OPC and ff14SB+TIP3P, and also between ff19SB+OPC and ff19SB+TIP3P (**Figure S12**). However when comparing between force fields, differences are observed.

As expected, the PDB distributions indicate that each of these amino acids samples unique features on the Ramachandran map. The ff14SB approach is clearly overly simplistic; when the same uncoupled Ala-based parameters are applied to all three amino acids (A, V and L), the peptides exhibit very similar φ/ψ sampling during MD, with the only apparent difference being slight changes to the population of the β basin (**Figure 6**). This result is consistent with the ff14SB potential energy maps (**Figure 3** and **Figure 5**) where only subtle differences in β basins are observed between Ala and Val. The ff14SB population maps also lack the diagonal shape of the α basin that is clearly seen in the PDB data (also apparent in the dipeptide QM data discussed above). In contrast, using amino-acid specific training against QM data with solvent polarization, the differences in Ramachandran maps are reproduced much better with ff19SB CMAPs. For instance in the PDB, Val and Leu both have a flatter β-ppII transition region than Ala, with Val preferring greater population in this transition region. Compared to Ala, Leu has a broader diagonal α basin extending into the positive ψ region; these differences are reproduced more faithfully with ff19SB than ff14SB. The relative insensitivity of ff14SB backbone sampling to amino acid identity also explains its poor ability in modeling sequence dependence as discussed in the Introduction. Overall, given the fact that PDB data were not used in ff19SB training, this agreement between ff19SB and PDB shows a remarkable improvement in reproducing sequence-dependent behavior obtained using physics-based training, and highlights that these trends can be recapitulated without problematic empirical fitting against PDB data.

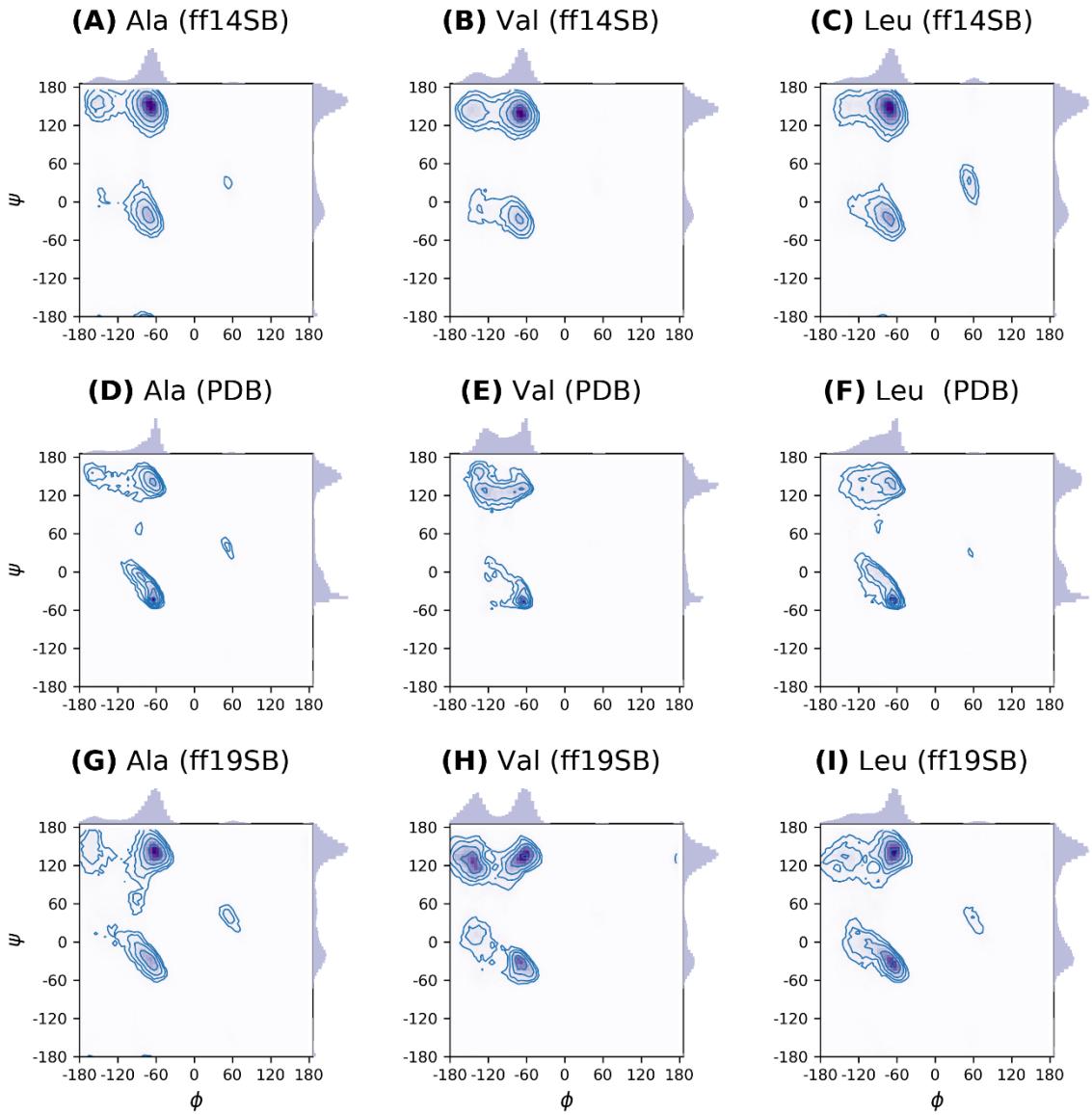

Figure 6. Ramachandran sampling shown for Ala, Val and Leu in dipeptide simulations with OPC water and ff14SB (A)-(C), in PDB (by Lovell et al.[42, 98]) (D)-(F), in dipeptide simulations at 300K with OPC water and ff19SB (G)-(I). Each contour line represents a doubling in population. Density is also shown as grids filled with light (no density) to dark (maximum density). Side histograms on each subplot represent independent distributions on φ and ψ. The MD simulations were run for a total of ~10 μs for all data shown.

Improved reproduction of NMR $^3J(HNHA)$ scalar couplings on blocked dipeptides

Another way to examine the ability of ff19SB to improve amino-acid specific behavior in solution is through quantitative comparison against NMR data probing backbone dihedrals, which have been reported[91] for each of the amino acids in a dipeptide form (except Pro which lacks HN). As explained (**Methods: CMAP fitting groups**), a total of 16 CMAPS were fit and then applied to 20 natural amino acids (also including alternate side chain protonation states) in ff19SB. We compared the performance of ff19SB and ff14SB by simulating blocked dipeptide systems (**Methods: Structure preparation & simulations**) in both OPC and TIP3P solvent models. We then calculated the $^3J(HNHA)$ from each MD trajectory based on the Karplus equation[89] and "Orig" parameter set[90] and quantified the agreement by calculating the χ^2 error following Best et al[15] and us[88]. This χ^2 error was used as an empirical target in ff14SB backbone training[10]. The χ^2 value quantifies the agreement between experimental and MD ensemble average J value(s), also taking into account the uncertainty of the theoretical model being used. In theory, smaller χ^2 errors correspond to better agreement between MD and experiment. However, χ^2 values below one only indicate that the error is smaller than the uncertainty of the model and do not necessarily indicate continued improvement vs. experiment. Further details of the calculations and precision estimates are provided in Methods (**Methods: NMR scalar coupling calculations**).

The calculated $^3J(HNHA)$ values for each amino acid, using four different combinations of FF (ff14SB and ff19SB) and water model (OPC and TIP3P), are provided in **Table S9**, with the χ^2 errors for OPC shown in **Figure 7** and TIP3P shown in **Figure S13**. Though we observed differences among force fields for the Ramachandran sampling maps, the χ^2 errors and actual $^3J(HNHA)$ values appear relatively insensitive to force field. For a given force field, neither

Ramachandran sampling maps nor the χ^2 errors and actual 3J(HNHA) values are sensitive to solvent model. For instance, for either ff14SB or ff19SB, the average χ^2 errors are similar and mostly below 0.5 for both OPC and TIP3P (**Figure 7** and **Figure S13**). In this respect, the performance of ff19SB is not significantly improved over ff14SB for dipeptide NMR data, as ff14SB already showed reasonable behavior with few amino acids having errors larger than 1.0 (His^+ and Cys) for both solvent models. In addition, the histograms of χ^2 errors are similar regardless of the force field and solvent model (**Figure S14**). Together with the fact that *$^3J(HNHA)$* in the Karplus calculation is sensitive only to the φ dihedral, this test seems insufficient to examine the specificity of parameters for different amino acids and the quality of parameters across the full Ramachandran space. However, this is a good indicator that the QM fitting is reasonable and ff19SB introduced no spurious outliers.

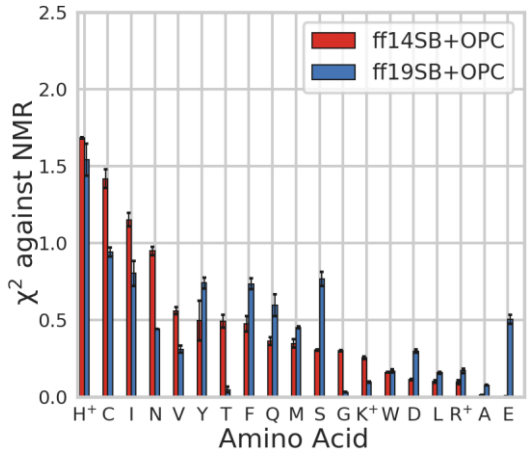

Figure 7. χ^2 errors in reproducing NMR *$^3J(HNHA)$* coupling data for all non-Pro amino acids (using single letter codes on X axis), with data for ff14SB+OPC (red) and ff19SB+OPC (blue). The MD simulations were run at 300K for a total of ~60 μs for all data shown.

As shown in **Figure 7**, ff19SB+OPC gave a slightly larger error for Glu, but since the pH used in the NMR experiment (4.9) was close to the Glu side chain pKa (~ 4.25), a simulation using either a protonated or deprotonated state of Glu may not adequately model the experimental ensemble. To address this ambiguity, we ran constant pH simulation (pH=4.9) on Glu dipeptide (**Methods: Constant pH simulation**), and obtained the carboxyl group protonated state ratio for each force field + solvent model combinations (**Table S12**). Next, we performed regular MD for both protonated and deprotonated Glu. The combined trajectory weighted by protonation state ratio (**Methods: Constant pH simulation**) was used so that our calculated χ^2 more accurately reflected the protonation states in the experiment.

For deprotonated Glu, the ppII region is the most populated in both ff14SB and ff19SB and the shape of energy basins are similar between ff14SB and ff19SB regardless of the solvent model (**Figure S12**). However, ff19SB samples the ppII basin extending farther towards $\varphi > -60°$ than ff14SB. This subtle change causes the $^3J(HNHA)$ to deviate significantly from experiment (χ^2 = 1.31±0.03). This shift, however, is much less pronounced in the protonated state MD with ff19SB (**Figure S12**), resulting in a much smaller χ^2 error of 0.031±0.01. Overall, the χ^2 value from the re-weighted population at pH 4.9 was calculated to be 0.50±0.03, indicating that the scalar coupling calculated with ff19SB is in reasonable agreement with experiment once the protonation state is taken into account.

We also performed constant pH simulation at pH=4.9 for Asp, obtaining the side chain carboxyl protonation ratio for different force field + solvent model combinations (**Table S12**). The χ^2 values from Asp simulation with deprotonated side chain and pH-weighted ensemble were calculated to be 0.01±0.01 and 0.30±0.01, respectively, with both indicating reasonable

agreement with experiment for ff19SB. In addition, for both Asp and Glu, in either ff14SB or ff19SB simulations, using TIP3P vs. OPC has little effect on the χ^2 results with average χ^2 errors all below 0.5.

In summary, both ff19SB and ff14SB provided reasonable results in reproducing NMR scalar coupling when using either OPC or TIP3P solvent, indicating that this test is relatively insensitive to the sampling differences that are apparent in the Ramachandran surfaces (**Figure S12**). It is encouraging, however, that ff14SB includes an empirical adjustment to improve agreement with the same type of NMR data as used here, while the QM-trained ff19SB achieves similar or better accuracy without empirical adjustment.

Accurate reproduction of Ala_5 NMR scalar couplings is maintained in ff19SB

We next tested ff19SB by simulating Ala_5 in both OPC and TIP3P solvents, and compared to ff14SB. A total of six NMR scalar couplings have been measured on this peptide[43]. Following Best et al.[15] and us[10, 88] previously, we calculated the scalar couplings from each MD trajectory as discussed above, and quantified the agreement between simulations and NMR by calculating the χ^2 error (**Methods: NMR scalar coupling calculations**). The NMR data, calculated scalar couplings for ff14SB and ff19SB in both OPC and TIP3P water and the systematic error σ[15, 88] used in χ^2 calculations are provided in **Table S10**, with the χ^2 errors in OPC shown in **Figure 8** and TIP3P shown in **Figure S15**. Overall, the average χ^2 errors are smaller than one regardless of force field and solvent model, indicating a reasonable reproduction of NMR data for ff14SB and ff19SB with both OPC and TIP3P. Specifically, ff19SB has smaller averaged χ^2 compared to ff14SB for both OPC (0.77±0.03 vs. 0.93±0.10) and TIP3P (0.77±0.03 vs. 0.88±0.09) solvent model. The measurement of $^3J(HNCA)$ is correlated with the ϕ dihedral as well as the ψ dihedral

of the preceding amino acid[43, 89]; this is the only coupling we examined that depends on two dihedrals instead of one. This Karplus correlation has the smallest σ among all of these scalar coupling types, making it more sensitive to error than other scalar coupling types. Even though the χ^2 value is large (**Figure 8** and **Figure S15**), the difference between simulation and NMR in actual *$^3J(HNCA)$* value is as small as 0.2 across all models, suggesting reasonable agreement between simulation and NMR across different models (**Table S10**).

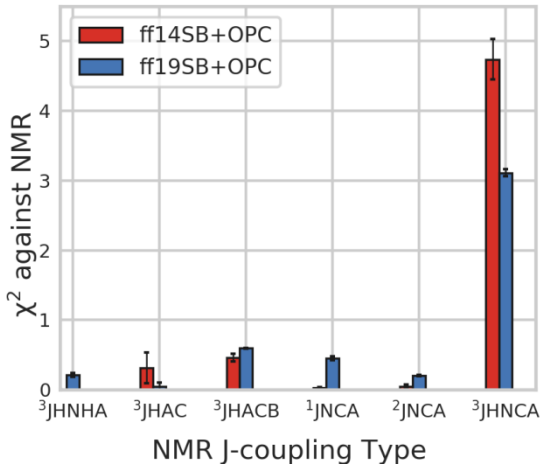

Figure 8. χ^2 errors in reproducing six NMR scalar coupling data for Ala_5, with data for ff14SB+OPC (red) and ff19SB+OPC (blue). The MD simulations were run at 300K for a total of ~3 μs.

Amino-acid specific helical propensities are significantly improved in ff19SB

Since the scalar coupling χ^2 analysis presented above was relatively insensitive to the updated residue-specific parameters, additional tests were performed to further validate the new model. The *$^3J(HNHA)$* analysis is only sensitive to the distribution for φ; thus, we calculated amino-acid

specific helical propensities to probe ψ dihedral sampling. We focus both on the absolute helical propensity in the force field as well as the ability to reproduce known differences between amino acids. We performed multiple MD simulations on model peptides with sequence Ace-A_4**X**A_4-NH_2 with varying **X**, and fit helical propensity parameters *w* through Lifson-Roig[86] theory implemented in a genetic algorithm (**Methods: Helical propensity**). Different from having three substitutions in Best et al.'s system[9], our model peptides only have a single substitution, as was done for the experimental system[39], to avoid possible interaction between the substitutions across turns of helix. The sensitivity to the peptide length was tested by comparing propensities calculated using A_4**X**A_4 and A_9**X**A_9 in ff14SB + GBneck2; calculated helical propensities for all amino acids with ff14SB + GBneck2 are highly correlated between A_4**X**A_4 and A_9**X**A_9 (**Figure S16**), justifying the use of the shorter peptide in the more computationally expensive explicit solvent simulations.

We also calculated the sensitivity of the results to the exact definition of the helical region of overall φ/ψ space (defined in **Table S6**) using ff14SB and ff19SB, in both OPC and TIP3P. The calculated helical propensities for each force field and solvent model show little sensitivity to theα basin definition, especially for ff19SB+OPC (**Figure S4**).

Helical propensities were calculated for A_4**X**A_4 with ff14SB and ff19SB, in TIP3P and OPC. The results of the MD simulations are compared to values based on experiments[39]. Data for ff14SB+TIP3P, ff14SB+OPC, ff19SB+TIP3P and ff19SB+OPC are shown in Histidine is a special case because the imidazole protonation state (δ, ε or both) is difficult to assign, and the reported experimental scales for 20 natural amino acids vary the most for His, with it being the least helical from one experimental scale but almost in the middle of the helicity from another[39, 103] For instance, Pace and Scholtz[103] summarized a helical propensity scale based on NMR

measurements of helix propensity from 11 systems, including both proteins and short peptides, at different pH values and temperatures. All helical propensities were reported in ΔΔG relative to Ala (0 kcal/mol, the most helical) and normalized by setting Gly=1 kcal/mol, the least helical. In that report, His exhibits a value of 0.61±0.11 (error bar calculated from 13 reported measurements) averaged across systems and protonation states (estimated based on experimental pH). Specifically, for neutral His, the helical propensity is 0.56±0.07 (uncertainty calculated from seven reported measurements), and for the protonated His^+, the helical propensity is 0.66±0.10 (uncertainty calculated from six reported measurements). This value is much lower (closer to Ala, meaning more helical) than several other amino acids including Asn, Thr, Cys and Asp. However, according to the NMR data[39] (reported as helical propensity *w* instead of ΔΔG), His is the least helical along with Gly (see **Table S7**). These NMR data are generally consistent with Pace and Scholtz except for His (**Figure S5**). Due to these uncertainties, we decided to remove His from the helical propensity comparisons in **Figure 9**. The helical propensity data (from both NMR and MD) including His are provided in **Table S7** and **Table S8**.

In general, ff14SB has difficulty reproducing the trend from NMR experiments regardless of solvent model. In TIP3P, Ala should be the most helical amino acid but is distinctly underestimated, while most other amino acids have significantly overestimated helical propensities, and the overall residue-specific correlation with NMR is poor at $R^2 = 0.38$ (**Figure 9A**). Although OPC is arguably a better water model[32] than TIP3P, combining it with ff14SB produces worse results than in TIP3P ($R^2 = 0.27$, **Figure 9B**), with helical propensities being underestimated for most amino acids. There is very little sequence dependence, with a slope of 0.49. The amino acids with negatively charged side chains (Asp and Glu) are outliers in both

solvent models for ff14SB. The correlation and sequence dependence furtherworsen when Asp and Glu are removed (R^2 = 0.18 and Slope = 0.22).

This poor correlation with experiment appears to be due to ff14SB rather than weaknesses in these solvent models; the correlation is significantly higher when comparing the helical propensities of ff14SB in two water models (OPC vs. TIP3P R^2 =0.84 as shown in **Figure S17**, with TIP3P giving ~70% higher helical propensities than in OPC). These results suggest that the ff14SB force field would be unable to reliably model quantitative changes to secondary structure or protein stability due to point mutations, despite its ability to successfully fold large proteins to near-native structures[37]. Protein folding tests are likely less sensitive to sequence-specific

energetics since the overall fold can be maintained even when a large fraction of the protein sequence is varied[104].

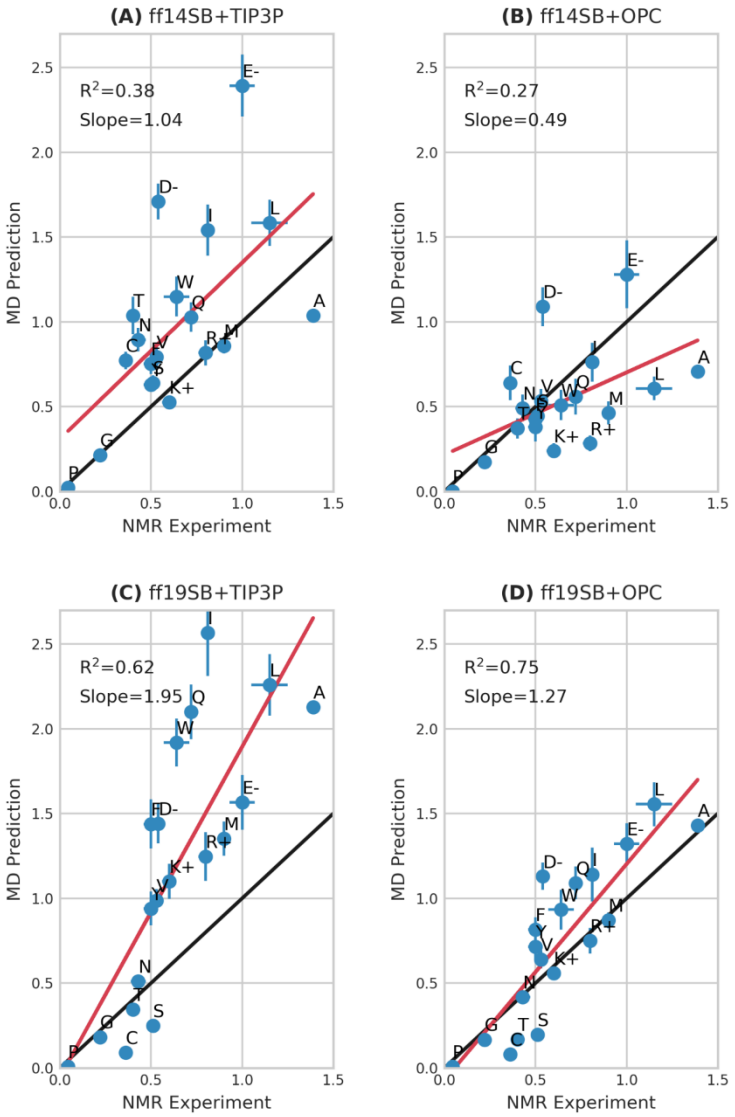

Figure 9. Correlation between helical propensities w from experiment[39] and simulations using (A) ff14SB+TIP3P, (B) ff14SB+OPC, (C) ff19SB+TIP3P and (D) ff19SB+OPC. Amino acids are indicated using single letter codes. Values on the X-axis represent the data based on NMR[39] and the reported standard deviations. Values on Y-axis represent the helical propensities fit against the combined trajectory (3.2 μs * 12), with error bars calculated via bootstrapping analysis. Black lines represent perfect agreement. Linear regression (red lines) was performed

against the data points, with R^2 and slope quantifying the goodness of fit. The MD simulations were run at 300K for a total of ~3225 μs.

Ideally, the ff19SB residue-specific training against QM data should improve modeling of sequence-dependent behavior and give improved correlation to experimental residue-specific differences. Consistent with this expectation, we find that using ff19SB+TIP3P reproduces the experimental trend much better than ff14SB+TIP3P (R^2 = 0.62 vs 0.38, respectively, **Figure 9C** vs. **9A**). However for ff19SB+TIP3P we also observe substantially higher sensitivity to amino acid variation than in experiment (slope = 1.95, **Figure 9C**). The source of this high slope and amplified sensitivity may be weaknesses in TIP3P (see **Introduction**), in particular the bias favoring compact structures like helices.

When ff19SB is combined with the better water model OPC (**Figure 9D**), the correlation between simulated and experimental helical propensities is further improved (R^2 = 0.75 vs. 0.62 in TIP3P) and the sensitivity to amino acid is also improved (slope = 1.27 vs. 1.95 in TIP3P). The sensitivity of the model still seems slightly overestimated, with slope modestly larger than unity. The remaining deviations from a perfect linear correlation may not be highly significant, since small disagreements also exist among various experimental measurements (**Figure S5**). In OPC, the helical propensity for Ala remains slightly too low with ff19SB, and Leu is similar to Ala within uncertainties (**Table S7**). Ser, Thr and Cys are all predicted to have helical propensity somewhat lower than experiment; all have short, polar side chains that could compete with backbone hydrogen bonding and reduce helical content. This will be investigated in more detail in the future.

These results show that ff19SB has significantly improved capability to differentiate amino acid properties and thus should have better predictive power for modeling sequence-specific

behavior, protein mutations, and also rational protein design which requires quantitative sequence-structure accuracy.

In addition to ff14SB and ff19SB, we also considered several other recent Amber-related force fields in combination with their recommended water model (**Table S8**). In ff15ipq[23]+SPC/E_b[24], Ala shows good agreement with experiment, but otherwise there is poor overall correlation and weak sensitivity among the remaining amino acids (R^2 = 0.26 and slope = 0.52, **Figure S18A**). In fb15[57]+fb3[61], Ala helical propensity is much lower than NMR, and the overall correlation is also poor (R^2 = 0.28 and slope = 0.74, **Figure S18B**). Best et al. reported helical propensity benchmarks for 20 amino acids, showing that the overall trend from experiments[39] was poorly reproduced by two force field + water combinations (ff03w[105]+TIP4P/2005[106] and ff99SB*[13a]+TIP3P[60]) with correlation coefficients R^2 being 0.01 and 0.22 respectively.[9] Therefore, they performed an empirical adjustment of a few amino acids, together with the updated parameters in the ILDN[107] variants of ff99SB*, to better match helix-coil transition data. They refit partial charges of Cα and side chain atoms on charged amino acids (D, E, K, R) while forcing the charges on amide N, H, C, O to have same values as all the other residues. The helical propensities[9] using these charge-refit residues were better correlated with experiment (ff99SB*_ILDN_Q + TIP3P, R^2 = 0.51 and slope = 0.68 for all amino acids, **Figure S19**) than the original ff03w and ff99SB*, but even with this empirical fitting the overall trend for the 20 amino acids is still notably worse than ff19SB+OPC (R^2 = 0.75 and slope =1.27, **Figure 9D** & **Figure S19**).

Evaluating helical content in the K19 peptide

In order to assess the ability of ff19SB to model α-helices in more complex systems, we employed the Ala-rich Baldwin-type[108] peptide K19[64] that was previously simulated[10] using ff14SB. Experimental measurements[64] on K19 using NMR chemical shift deviations (CSDs) suggest that the fraction helix at 300 K of four central residues and two residues near the C-terminus are ~0.38 and ~0.17, respectively (**Figure 10**). Simulations with ff14SB+TIP3P exhibited an average 0.30±0.05 (central four) and 0.19±0.03 (two near C terminus) fraction helix, in close agreement with our previously reported[10] value of 0.30±0.05 and 0.20±0.04 using the same force field and solvent model. Both values are in good agreement with the experiment, likely reflecting the inclusion of K19 data generated using TIP3P in the empirical adjustment of ff14SB backbone parameters.

In order to better separate the accuracy of the solute force field from that of the solvent model, we ask: does the good match come from a good modeling of protein and water separately, or from training-based error cancellation between the force field and solvent model? As shown in **Figure 10**, after substituting TIP3P with a better model for water (OPC), ff14SB MD resulted in significantly reduced helicity, with 0.08±0.02 (central four) and ~0.08±0.01 (two near C terminus) helical content for the 6 measured residues. Given OPC's excellent agreement with water properties, the worsened agreement with experiment for K19 supports a fortuitous cancellation of error in the combination of ff14SB+TIP3P. Since overly weak solvent-solute dispersion in TIP3P[58a, 58c] may introduce a bias in favor of compact structures, it seems reasonable that this bias may also enhance helical content to maintain hydrogen bonding in compact states. This hypothesis is supported by data in **Figure S20**, which shows an inverse correlation between helical content and radius of gyration of K19, indicating that more compact structures tend to be more helical, and also A4**X**A4 data in **Figure 9**, which shows a dramatic

increase in helical propensities when combining ff19SB with TIP3P vs. OPC. We conclude that an inherent underestimation of helicity is present in ff14SB, which is (inexactly) compensated by an increase in helical content driven by the TIP3P bias toward overly compact structures.

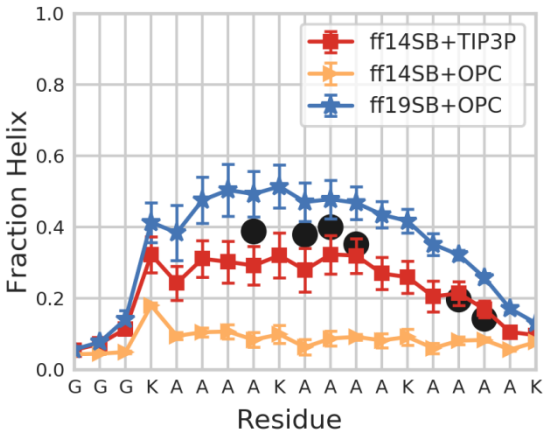

Figure 10. The fraction helix of each amino acid in K19 sampled in simulations using ff14SB+TIP3P (red), ff14SB+OPC (yellow) and ff19SB+OPC (blue). Uncertainties reflect the standard deviation of 10 independent runs. The black dots represent values reported in NMR experiments at 300 K[64]. The MD simulations were run at 300K for a total of ~96 μs.

Simulation of K19 with ff19SB+OPC resulted in modestly increased helical content vs. ff14SB+TIP3P, with 0.48±0.05 (central four) and ~0.29±0.01 (two near C terminus) average helicity. These values are also somewhat higher than those from experiment (~0.38 and ~0.17 respectively), but the deviation in MD corresponds to an error of only 0.24 and 0.35 kcal/mol free energy, respectively. Furthermore, uncertainties were not reported for the NMR-based data for K19, and ff19SB+OPC is in quantitative agreement with experiment for helical propensities for Lys and Ala that make up the majority of K19 (**Figure 9**). The simulations also agree with the trend from experiments, with the helical content falling off towards the C-terminus, with the

two measured Ala in this region being less helical than the central four. Overall, we conclude that the QM-based ff19SB is in reasonable agreement with experiment when combined with an accurate solvent model, while ff14SB performs poorly with the same solvent model and relies on cancellation of error with the less accurate TIP3P model in order to reproduce the helical content of this alanine-based peptide.

β-hairpin stability

We next tested whether the improvements in modeling helical content with ff19SB (and perhaps a slight overestimation of helical content) were obtained at the cost of less accurate performance on β systems. Following our previous work[10] with ff14SB, we used CLN025[65], an engineered fast-folding hairpin that is a thermally optimized variant of Chignolin[65]. CLN025 contains four aromatic side chains, including three Tyr and one Trp. This system presents a challenge due to the relatively slow folding of β-sheets compared to the helical systems (though T-jump IR experiments[66] estimate a 100-ns folding time for CLN025). Because of the computational cost in obtaining highly precise estimates of β hairpin population in MD simulations with explicit water, we limit our testing here to a qualitative view of whether ff19SB's improved helical propensity prediction may compromise β stability. We again tested ff14SB with TIP3P and OPC, and ff19SB with OPC.

We performed four MD runs, each of 7 μs in length, at 300K starting from the NMR structure (two runs) and a fully extended structure (two runs) (56 μs total for all ff+water combinations). As measured by backbone RMSD against the NMR structure (PDBID: 2RVD[65]), folding was reversible in every simulation using each of the three combinations of the force field + solvent model (**Figure 11**). The histogram of RMSD values shows that both ff14SB+TIP3P and

ff19SB+OPC predominantly sample the NMR structure (**Figure S2**). The average fraction of native population (±standard deviation) across all MD runs for ff14SB+TIP3P, ff14SB+OPC and ff19SB+OPC are 0.75±0.23, 0.33±0.19 and 0.50±0.17, respectively, compared to the experimental estimate[65] of 0.9 based on CD spectra. These populations suggest that ff14SB+TIP3P might stabilize the β-hairpin to a greater extent than the other combinations, but the differences are within the uncertainties of the populations. It is interesting that with ff14SB, MD in TIP3P appears to prefer more β-hairpin structure than with OPC. The same preference for the native structure in TIP3P was seen with K19, perhaps indicating that the weaker solute-solvent dispersion in TIP3P generally stabilizes compact structure (such as native folds) consistent with previous studies[29-30, 58], rather than a specific secondary helical bias such as the K19 stability increase discussed above.

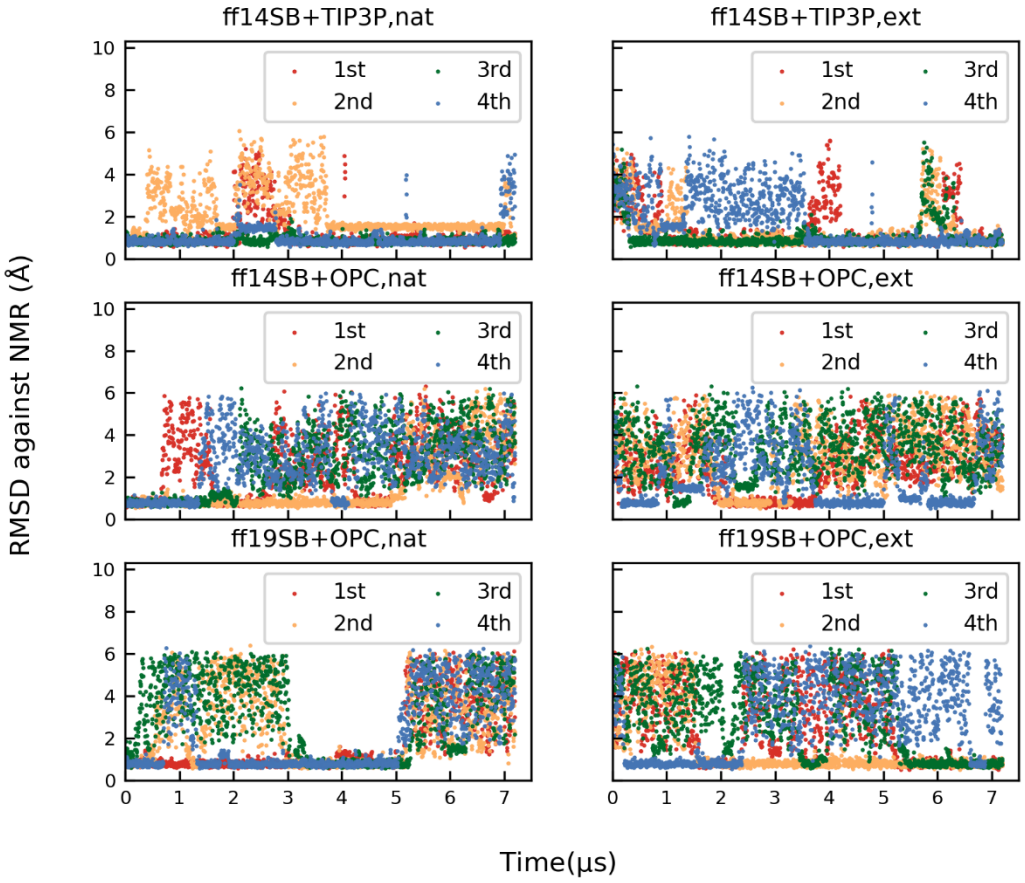

Figure 11. Backbone RMSD to the NMR structure (PDBID: 2RVD[65]) vs. time for the four extended (ext) and four native (nat) runs of CLN025 with ff14SB+TIP3P, ff14SB+OPC and ff19SB+OPC. The MD simulations were run at 300K for a total of ~172 μs.

High quality backbone dynamics vs. NMR is maintained with ff19SB

NMR S^2 order parameters reflect the internal protein dynamics that are helpful to validate MD trajectories. These internal motions need to be separated from global tumbling on time scale of pico to nanosecond. Therefore, a choice for the window size of the MD trajectory needs to be made over which S^2 values are computed and averaged, which remains challenging[95-96]. As

reported in our previous work, ff14SB+TIP3P maintained ff99SB's overall good reproduction of NMR S^2 order parameters.[10] Here, we also evaluated the ability of ff19SB to recapitulate local dynamics in well-folded proteins. Interestingly, ff19SB+OPC samples structures with even lower RMSD against native crystal structure than either ff14SB+TIP3P or ff14SB+OPC (**Figure S21-S23**). As shown in **Figure 12**, the NMR data were reasonably reproduced by the different force field + solvent model combinations, with average absolute difference between NMR S^2 and calculated S^2 over all amino acids close to 0.04. The overall differences were not statistically significant; however we note some instances where all three force field + solvent models deviate from experiment and also some instances where ff19SB results are in worse agreement with experiment than is ff14SB. These residues typically have overestimated flexibility in MD as compared to NMR for Gly (smaller S^2 in MD). I Rendered crystal structures of the three proteins are provided in **Figure S24**, highlighting the locations of residues that are discussed below. All NMR and MD data are provided in **Table S13-S15**.

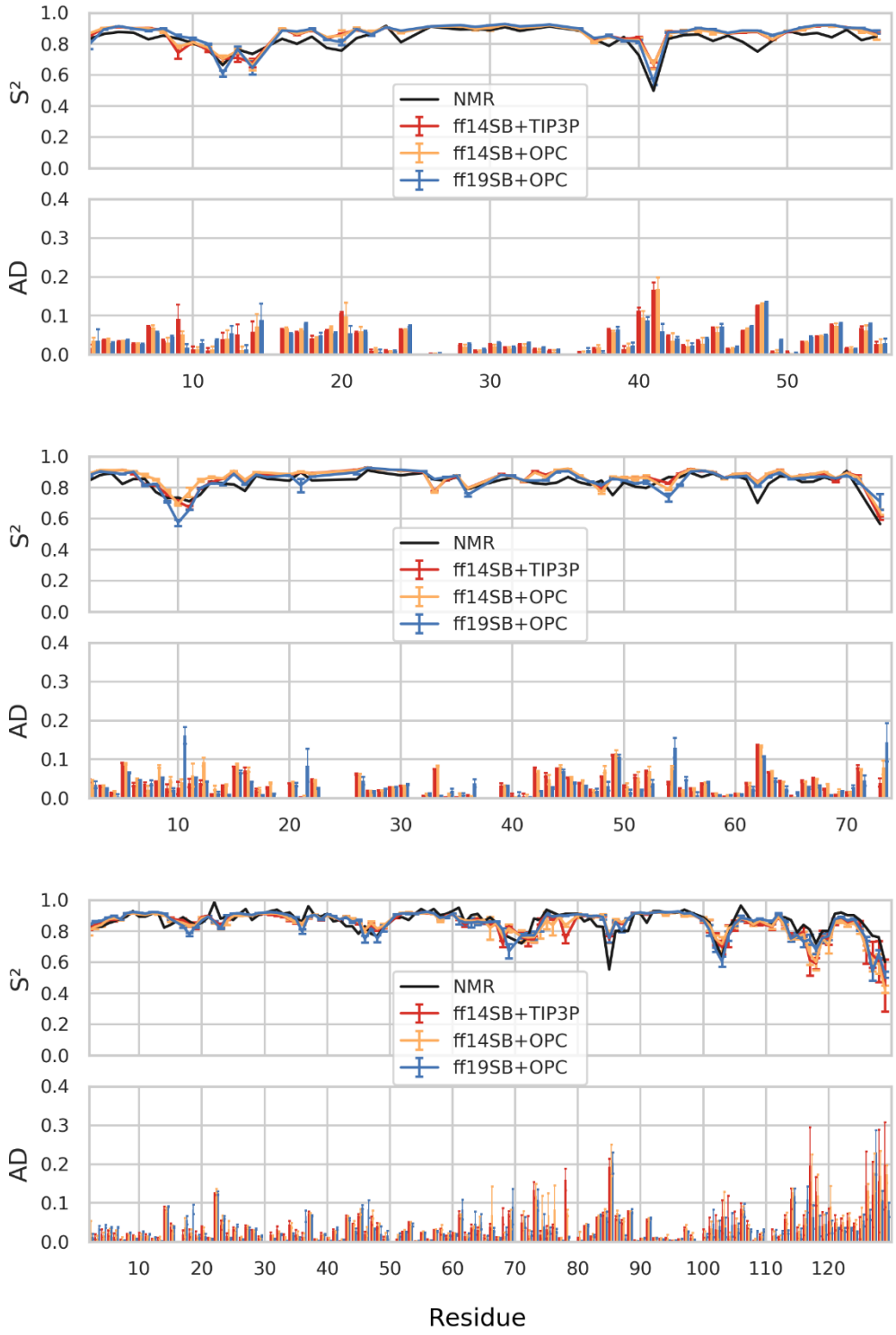

Figure 12. Per-residue order parameters (S^2) from NMR compared to simulations at 300K using ff14SB+TIP3P (red), ff14SB+OPC (yellow) and ff19SB+OPC (blue) of (top) GB3[109], (middle)

Ubiquitin[110] and (bottom) Lysozyme[111]. AD is the absolute difference between NMR and MD simulation. For each subplot, error bars represent the standard deviation from four independent runs. Some residues are missing experimental values as indicated in the original NMR papers[109-111]. The MD simulations were run at 300K for a total of ~1.8 μs.

In GB3, Gly14 was reported to have a high S^2 (0.74) using NMR, likely due to its intermolecular hydrogen bond in a β-sheet secondary structure. However, Gly14 is similarly more flexible with ff19SB+OPC (0.65±0.04), ff14SB+TIP3P (0.68±0.03) and ff14SB+OPC (0.66±0.03). This may not reflect problems in ff19SB Gly parameters since this trend is reversed for Gly41 in the loop region connecting a β-strand to an α-helix. S^2 from NMR is quite low for Gly41 (0.50) due to loop flexibility, and this flexibility is reproduced much better with ff19SB+OPC (0.56±0.04) than ff14SB+TIP3P (0.66±0.02) and ff14SB+OPC (0.67±0.03).

In Ubiquitin, Gly10 flexibility is overestimated in ff19SB+OPC (0.55±0.02), but not in ff14SB+TIP3P (0.72±0.02) and ff14SB+OPC (0.69±0.01) compared to NMR (0.73). Except for the slightly worsened performance on Gly10, ff19SB+OPC yields the best overall agreement with NMR compared to ff14SB with either TIP3P or OPC solvent model.

In Lysozyme, Ser85 lies in a loop region connecting two α helices, and is overly rigid with all three simulation models (~0.75 in MD vs 0.55 in NMR). However, Ser85 backbone (φ/ψ) and side chain (χ1/χ2) sampling in all three force field + solvent model combinations reproduces that seen in the crystal structure.

In spite of subtle disagreements with NMR in both models, we concluded that ff19SB generally maintained the overall performance of ff14SB and ff99SB in order parameter reproduction, with a few outliers that do not appear to follow any systematic trend that could be attributed to the CMAPs.

Conclusions

In the updated ff19SB (with the "SB" models indicating Stony Brook) protein force field presented here, we have developed new backbone dihedral parameters with amino-acid specific CMAP functions. We trained the parameters to match solution phase QM data using full 2D φ/ψ scans, instead of the gas-phase minima used for training uncoupled φ and ψ cosine terms in ff99SB. Use of energies calculated from QM in solution provides better consistency with the pre-polarized partial atomic charges used by the MM model, as compared to gas-phase energies that were used previously. Fitting of dihedral corrections against QM in solution also allows the model to incorporate (to some extent) conformation-dependent polarization energy that is not present explicitly in a fixed-charge MM model such as the one used here.

A total of 16 CMAPs were fit, with applicability to all amino acids using a grouping approach based on side chain size, branching and polarity. Leu was used as a general model for other amino acids, in contrast to Ala that has traditionally been used as a protein backbone model. We also investigated whether CMAP functions fit using a single side chain rotamer could remain accurate for other rotamer states, and found good transferability as measured by the ability of the model to reproduce rotamer-dependent differences in Ramachandran space QM energetics and PDB-based statistics.

One possible weakness to our approach was the use of simple implicit water models during training, such as the GB model in the MM component. Older GB models exhibit secondary structure biases for longer peptides[112], but here we have used our GBneck2 model[62] that much more accurately reproduces secondary structure preferences. Furthermore, we have shown that the solvation energy of dipeptides (which we used here for the CMAP training in GB) is largely

insensitive to specific GB model used[102]. Nevertheless, our use of GB during training could be a limitation, which is one reason we carried out extensive testing here using a variety of fully explicit water models.

We performed a total of ~5 milliseconds MD simulations in explicit solvent to extensively validate ff19SB against experiments. The results show that our new FF more accurately reproduces amino-acid specific NMR properties such as scalar coupling and helical propensity, as well as structure and stability of a Baldwin-type helical peptide and a small hairpin. While the balance of secondary structure seems reasonable in our helix and hairpin runs, the results may suggest a slight overstabilization of helix vs. sheet. More quantitative testing via longer simulations on a larger variety of systems is needed to see if this is a general trend in the model. Folded proteins show good agreement with NMR S^2 order parameters, and modestly improved RMSD values as compared to ff14SB.

We make the important observation that the performance of the QM-based ff19SB model improves as the quality of the water model is improved (going from TIP3P to OPC), suggesting lack of fortuitous cancellation of error with a particular water model, and that the water model is likely the limiting factor in these comparisons of ff19SB to experiment. Behavior with more water models will be tested in the future to see if this is a general trend. Currently, our best results are obtained using ff19SB with OPC water, and we recommend that combination. Since ff19SB was not fit using OPC, however, there is no reason to expect that better performance cannot be obtained using other models not tested here. Biomolecular force fields such as ff19SB that are not tied to a specific water model through empirical adjustment will be in a stronger position to take advantage of future, better-quality water models, ideally with independent training of these components. In contrast, use of a better model for water does not lead to

improved match with experiment for ff14SB that was empirically trained using TIP3P, supporting that both a good water model and good protein force field are needed for an accurate simulation. We also conclude that weaker solute-solvent dispersion in TIP3P not only leads to overly compact unfolded states as has been reported previously, but also overstabilizes both secondary structure elements as compared to OPC.

If water models can be sufficiently improved, there is in principle no need for specialized "IDP" force fields, as suggested in recent work[29] by Robustelli et al. Our belief is that physics-based protein FFs trained against short peptides should be quite capable of modeling IDPs and unfolded ensembles, which are more similar to the peptide training data than are folded proteins. Amber's OPC 4-point water model not only better reproduces liquid water properties as compared to most other models[32], but IDP simulations with OPC result in much less compact ensembles as compared to simulations using the same FFs in older water models.[33] This provides additional evidence that the current problems with modeling IDPs are likely to be related to the water models, and further improvement of physics-based protein FFs is warranted, independent of water model development going on in parallel. While the studies here of flexible peptides using ff19SB+OPC are promising, future studies using this combination for IDPs will be carried out in the future.

ASSOCIATED CONTENT

Supporting Information. Additional files for ff19SB implementation test and PDB statistical analysis. Additional information about MD equilibration, protein system details, information on dihedral restraints, modified force field parameter files, CMAP fitting group, refitted partial charges, secondary definition, protonation state ratio, NMR J couplings from experiment and

MD, NMR helical propensities from experiment and MD, NMR S^2 parameters from experiment and MD, initial structures of K19 simulations, REE vs. CPU for QM methods, distribution of bootstrapped resampling, energy maps from QM, ff14SB and ff19SB, φ/ψ distribution on dipeptides, helical propensities from more force field and water model, RMSD vs time and histogram on RMSD for GB3, ubiquitin and lysozyme.

The following files are available free of charge.

Input and output files for validation of ff19SB implementations

PDB IDs and the residue numbers for statistical analysis (pdbid_residue.csv)

AUTHOR INFORMATION

Corresponding Author

*E-mail: carlos.simmerling@stonybrook.edu

Phone: +1 (631) 632-5424. Fax: +1 (631) 632-5405

Present Address

Dr. Angela N. Migues is currently at the State of University of New York College at Oneonta.

Author Contributions

Chuan Tian performed the force field parameterization and validation analysis under the guidance of Dr. Carlos Simmerling and Dr. Qin Wu. Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, Dr. He Huang, Dr. Angela N. Migues, John Bickel, Yuzhang Wang and Jorge

Pincay contributed to the project design, data analysis and interpretation. All authors have given approval to the final version of the manuscript.

Funding Sources

The authors acknowledge funding from the National Science Foundation (1665159), Blue Waters computing resources funded by National Science Foundation (Awards OCI-0725070 and ACI-1238993). Dr. Angela N. Migues was supported by the National Institute of General Medical Sciences of the National Institutes of Health (K12GM102778). This research used resources of the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science Facility, and the Scientific Data and Computing Center, a component of the Computational Science Initiative, at Brookhaven National Laboratory under Contract No. DE-SC0012704. The work was also supported by the Laufer Center at Stony Brook University.

ACKNOWLEDGEMENT

The authors gratefully acknowledge helpful suggestions from Dr. Daniel Raleigh.

ABBREVIATIONS

CD, circular dichroism; CSD, chemical shift deviation(s); CMAP, correction map; ESP, electrostatic potential; ff14SB, force field 2014 Stony Brook; ff19SB, force field 2019 Stony Brook; GB3, third immunoglobin-binding domain of protein G; HF, Hartree−Fock; HEWL, lyozyme; IDP, intrinsic disordered protein; ILDN, isoleucine, leucine, aspartate, and asparagine; iRED, isotropic reorientational eigenmode dynamics; K19, ACE-G3$(KAAAA)_3$K-NH2; MM, molecular mechanics; MD, molecular dynamics; MP2, Møller−Plesset Perturbation Theory of

the Second Order; NMR, nuclear magnetic resonance; Orig, “original” Karplus parameters; PDB, protein data bank; ppII, polyproline helix, type II; QM, quantum mechanics; REMD, replica exchange MD; RMSD, root mean squared deviations; Ubq, ubiquitin.

REFERENCES

1. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E., How Fast-Folding Proteins Fold. *Science* **2011,** *334* (6055), 517-520.
2. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society* **2015,** *137* (7), 2695-2703.
3. (a) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *Journal of chemical theory and computation* **2005,** *1* (1), 153-168; (b) Lamoureux, G.; MacKerell, A. D.; Roux, B., A simple polarizable model of water based on classical Drude oscillators. *J Chem Phys* **2003,** *119* (10), 5185-5197; (c) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T., Current Status of the AMOEBA Polarizable Force Field. *Journal of Physical Chemistry B* **2010,** *114* (8), 2549-2564; (d) Lopes, P. E. M.; Zhu, X.; Lau, A.; Roux, B.; MacKerell, A. D., Development of the Charmm Polarizable Force Field for Polypeptides Based on Drude Oscillators. *Biophys J* **2011,** *100* (3), 612-612.
4. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005,** *26* (16), 1668-88.
5. 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-Structure Function and Bioinformatics* **2006,** *65* (3), 712-725.
6. (a) 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. *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197; (b) Wang, J. M.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry* **2000,** *21* (12), 1049-1074.
7. (a) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C., Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *Journal of computational*

chemistry **2003,** *24* (1), 21-31; (b) Garcia, A. E.; Sanbonmatsu, K. Y., alpha-Helical stabilization by side chain shielding of backbone hydrogen bonds. *P Natl Acad Sci USA* **2002,** *99* (5), 2782-2787; (c) Kamiya, N.; Higo, J.; Nakamura, H., Conformational transition states of a beta-hairpin peptide between the ordered and disordered conformations in explicit water. *Protein Science* **2002,** *11* (10), 2297-2307; (d) Higo, J.; Ito, N.; Kuroda, M.; Ono, S.; Nakajima, N.; Nakamura, H., Energy landscape of a peptide consisting of alpha-helix, 3(10)-helix, beta-turn, beta-hairpin, and other disordered conformations. *Protein Science* **2001,** *10* (6), 1160-1171; (e) Ono, S.; Nakajima, N.; Higo, J.; Nakamura, H., Peptide free-energy profile is strongly dependent on the force field: Comparison of C96 and Amber95. *Journal of computational chemistry* **2000,** *21* (9), 748-762; (f) Wang, L.; Duan, Y.; Shortle, R.; Imperiali, B.; Kollman, P. A., Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures. *Protein Science* **1999,** *8* (6), 1292-1304.

8. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A., Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *Journal of the American Chemical Society* **1997,** *119* (25), 5908-5920.

9. Best, R. B.; de Sancho, D.; Mittal, J., Residue-Specific alpha-Helix Propensities from Molecular Simulation. *Biophys J* **2012,** *102* (6), 1462-1467.

10. 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. *Journal of chemical theory and computation* **2015,** *11* (8), 3696-713.

11. Janowski, P. A.; Liu, C. M.; Deckman, J.; Case, D. A., Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. *Protein Science* **2016,** *25* (1), 87-102.

12. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell Jr., A. D., Optimization of the Additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of chemical theory and computation* **2012,** *8* (9), 3257-3273.

13. (a) Best, R. B.; Hummer, G., Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *Journal of Physical Chemistry B* **2009,** *113* (26), 9004-9015; (b) 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 phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *Journal of chemical theory and computation* **2012,** *8* (9), 3257-3273.

14. Best, R. B.; Zheng, W.; Mittal, J., Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *Journal of chemical theory and computation* **2014,** *10* (11), 5113-5124.

15. Best, R. B.; Buchete, N. V.; Hummer, G., Are current molecular dynamics force fields too helical? *Biophys J* **2008,** *95* (1), L7-L9.

16. Georgoulia, P. S.; Glykos, N. M., Using J-Coupling Constants for Force Field Validation: Application to Hepta-alanine. *Journal of Physical Chemistry B* **2011,** *115* (51), 15221-15227.

17. Creamer, T. P.; Rose, G. D., Side-Chain Entropy Opposes Alpha-Helix Formation but Rationalizes Experimentally Determined Helix-Forming Propensities. *P Natl Acad Sci USA* **1992,** *89* (13), 5937-5941.

18. Shell, M. S.; Ritterson, R.; Dill, K. A., A test on peptide stability of Amber force fields with implicit solvation. *Journal of Physical Chemistry B* **2008,** *112* (22), 6878-6886.

19. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; 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. *Journal of Physical Chemistry B* **1998,** *102* (18), 3586-3616.
20. Chen, J.; Im, W.; Brooks, C. L., Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *Journal of the American Chemical Society* **2006,** *128* (11), 3728-3736.
21. Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A., ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *Journal of chemical theory and computation* **2014,** *10* (10), 4515-4534.
22. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* **2004,** *120* (20), 9665-9678.
23. 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. *Journal of chemical theory and computation* **2016,** *12* (8), 3926-3947.
24. Takemura, K.; Kitao, A., Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. *Journal of Physical Chemistry B* **2012,** *116* (22), 6279-6287.
25. Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H., Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *Journal of chemical theory and computation* **2015,** *11* (11), 5513-5524.
26. 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. *Current opinion in structural biology* **2014,** *24*, 98-105.
27. Palazzesi, F.; Prakash, M. K.; Bonomi, M.; Barducci, A., Accuracy of Current All-Atom Force-Fields in Modeling Protein Disordered States. *Journal of Chemical Theory and Computation* **2015,** *11* (1), 2-7.
28. (a) Jiang, F.; Zhou, C. Y.; Wu, Y. D., Residue-Specific Force Field Based on the Protein Coil Library. RSFF1: Modification of OPLS-AA/L. *Journal of Physical Chemistry B* **2014,** *118* (25), 6983-6998; (b) Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H. F., New force field on modeling intrinsically disordered proteins. *Chemical biology & drug design* **2014,** *84* (3), 253-69; (c) Song, D.; Wang, W.; Ye, W.; Ji, D.; Luo, R.; Chen, H. F., ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. *Chemical biology & drug design* **2016**.
29. Robustelli, P.; Piana, S.; Shaw, D. E., Developing a molecular dynamics force field for both folded and disordered protein states. *P Natl Acad Sci USA* **2018,** *115* (21), E4758-E4766.
30. (a) Henriques, J.; Cragnell, C.; Skepo, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *Journal of chemical theory and computation* **2015,** *11* (7), 3420-3431; (b) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated

Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B* **2015,** *119* (16), 5113-5123.
31. Petrov, D.; Zagrovic, B., Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *PLoS Comput Biol* **2014,** *10* (5), e1003638.
32. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J Phys Chem Lett* **2014,** *5* (21), 3863-3871.
33. Shabane, P. S.; Izadi, S.; Onufriev, A. V., General Purpose Water Model Can Improve Atomistic Simulations of Intrinsically Disordered Proteins. *Journal of chemical theory and computation* **2019,** *15* (4), 2620-2634.
34. 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. *Journal of chemical theory and computation* **2012,** *8* (4), 1409-1414.
35. Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A., Advances in free-energy-based simulations of protein folding and ligand binding. *Current opinion in structural biology* **2016,** *36*, 25-31.
36. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossvary, I.; Klepeis, J. L.; Layman, T.; Mcleavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y. B.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C., Anton, a special-purpose machine for molecular dynamics simulation. *Commun Acm* **2008,** *51* (7), 91-97.
37. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C., Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *Journal of the American Chemical Society* **2014,** *136* (40), 13959-13962.
38. Perez, A.; MacCallum, J. L.; Brini, E.; Simmerling, C.; Dill, K. A., Grid-Based Backbone Correction to the ff12SB Protein Force Field for Implicit-Solvent Simulations. *Journal of chemical theory and computation* **2015,** *11* (10), 4770-4779.
39. Moreau, R. J.; Schubert, C. R.; Nasr, K. A.; Torok, M.; Miller, J. S.; Kennedy, R. J.; Kemp, D. S., Context-independent, temperature-dependent helical propensities for amino acid residues. *Journal of the American Chemical Society* **2009,** *131* (36), 13107-16.
40. (a) 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. *Journal of computational chemistry* **2004,** *25* (11), 1400-1415; (b) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017,** *14* (1), 71-73.
41. Cornish, V. W.; Kaplan, M. I.; Veenstra, D. L.; Kollman, P. A.; Schultz, P. G., Stabilizing and Destabilizing Effects of Placing .beta.-Branched Amino Acids in Protein .alpha.-Helixes. *Biochemistry-Us* **1994,** *33* (40), 12022-12031.
42. Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C., The penultimate rotamer library. *Proteins* **2000,** *40* (3), 389-408.
43. 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. *Journal of the American Chemical Society* **2007,** *129* (5), 1179-89.

44. (a) Zhou, C. Y.; Jiang, F.; Wu, Y. D., Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of Amber ff99SB. *Journal of Physical Chemistry B* **2015,** *119* (3), 1035-1047; (b) Kang, W.; Jiang, F.; Wu, Y. D., Universal Implementation of a Residue-Specific Force Field Based on CMAP Potentials and Free Energy Decomposition. *Journal of chemical theory and computation* **2018,** *14* (8), 4474-4486.
45. Song, D.; Luo, R.; Chen, H. F., The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *J Chem Inf Model* **2017,** *57* (5), 1166-1178.
46. Feller, S. E.; MacKerell, A. D., An improved empirical potential energy function for molecular simulations of phospholipids. *Journal of Physical Chemistry B* **2000,** *104* (31), 7510-7515.
47. (a) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *Journal of the American Chemical Society* **1984,** *106* (3), 765-784; (b) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *Journal of computational chemistry* **1986,** *7* (2), 230-252.
48. (a) 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 (vol 117, pg 5179, 1995). *Journal of the American Chemical Society* **1996,** *118* (9), 2309-2309; (b) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J Phys Chem-Us* **1993,** *97* (40), 10269-10280.
49. Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophys J* **2006,** *90* (4), L36-L38.
50. Wang, L.-P.; Martinez, T. J.; Pande, V. S., Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *The Journal of Physical Chemistry Letters* **2014,** *5* (11), 1885-1891.
51. Mackerell, A. D., Empirical force fields for biological macromolecules: Overview and issues. *Journal of computational chemistry* **2004,** *25* (13), 1584-1604.
52. 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. Comp. Chem.* **2003,** *24* (16), 1999-2012.
53. Zgarbova, M.; Luque, F. J.; Sponer, J.; Otyepka, M.; Jurecka, P., A Novel Approach for Deriving Force Field Torsion Angle Parameters Accounting for Conformation-Dependent Solvation Effects. *Journal of chemical theory and computation* **2012,** *8* (9), 3232-3242.
54. Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A., Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. *Journal of Physical Chemistry B* **2013,** *117* (8), 2328-2338.
55. (a) Zgarbova, M.; Otyepka, M.; Banas, P.; Luque, F. J.; Cheatham, T. E.; Sponer, J.; Jurecka, P., Refinement of force field torsion parameters for nucleic acids based on inclusion of conformation-dependent solvation effects. *J Biomol Struct Dyn* **2013,** *31*, 70-70; (b) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurečka, P.,

Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of chemical theory and computation* **2011,** *7* (9), 2886-2902.
56. Zgarbova, M.; Rosnik, A. M.; Luque, F. J.; Curutchet, C.; Jurecka, P., Transferability and additivity of dihedral parameters in polarizable and nonpolarizable empirical force fields. *Journal of computational chemistry* **2015,** *36* (25), 1874-84.
57. 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. *Journal of Physical Chemistry B* **2017,** *121* (16), 4023-4039.
58. (a) Best, R. B.; Zheng, W. W.; Mittal, J., Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *Journal of chemical theory and computation* **2014,** *10* (11), 5113-5124; (b) Petrov, D.; Zagrovic, B., Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *Plos Comput Biol* **2014,** *10* (5); (c) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *Journal of Physical Chemistry B* **2015,** *119* (16), 5113-5123.
59. 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* (1), 235-242.
60. 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* (2), 926-935.
61. Wang, L. P.; Martinez, T. J.; Pande, V. S., Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J Phys Chem Lett* **2014,** *5* (11), 1885-1891.
62. Nguyen, H.; Roe, D. R.; Simmerling, C., Improved Generalized Born Solvent Model Parameters for Protein Simulations. *Journal of chemical theory and computation* **2013,** *9* (4), 2020-2034.
63. Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A., Generalized Born model with a simple, robust molecular volume correction. *Journal of chemical theory and computation* **2007,** *3* (1), 156-169.
64. Song, K.; Stewart, J. M.; Fesinmeyer, R. M.; Andersen, N. H.; Simmerling, C., Structural insights for designed alanine-rich helices: Comparing NMR helicity measures and conformational ensembles from molecular dynamics simulation. *Biopolymers* **2008,** *89* (9), 747-760.
65. Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K., Crystal Structure of a Ten-Amino Acid Protein. *Journal of the American Chemical Society* **2008,** *130* (46), 15327-15331.
66. Davis, C. M.; Xiao, S. F.; Raeigh, D. P.; Dyer, R. B., Raising the Speed Limit for beta-Hairpin Formation. *Journal of the American Chemical Society* **2012,** *134* (35), 14476-14482.
67. Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A., Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *Journal of the American Chemical Society* **2003,** *125* (30), 9179-9191.
68. Vijaykumar, S.; Bugg, C. E.; Cook, W. J., Structure of Ubiquitin Refined at 1.8 a Resolution. *J Mol Biol* **1987,** *194* (3), 531-544.

69. Young, A. C. M.; Dewan, J. C.; Nave, C.; Tilton, R. F., Comparison of Radiation-Induced Decay and Structure Refinement from X-Ray Data Collected from Lysozyme Crystals at Low and Ambient-Temperatures. *J Appl Crystallogr* **1993,** *26*, 309-319.
70. D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,; A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C.; Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I.; Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails,; R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman, Amber 2016. *University of California, San Francisco*.
71. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of computational chemistry* **2004,** *25* (9), 1157-1174.
72. Weiser, J.; Shenkin, P. S.; Still, W. C., Approximate solvent-accessible surface areas from tetrahedrally directed neighbor densities. *Biopolymers* **1999,** *50* (4), 373-80.
73. Lovell, S. C.; Davis, I. W.; Adrendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics* **2003,** *50* (3), 437-450.
74. Press, W. H.; Numerical Recipes Software (Firm) *Numerical recipes*, Code CD-ROM v 2.06.; Cambridge University Press,: Cambridge, 1996.
75. Crowley, M. F.; Williamson, M. J.; Walker, R. C., CHAmber: Comprehensive Support for CHARMM Force Fields Within the Amber Software. *Int J Quantum Chem* **2009,** *109* (15), 3767-3772.
76. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The journal of physical chemistry. B* **2009,** *113* (18), 6378-96.
77. Scalmani, G.; Frisch, M. J., Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J Chem Phys* **2010,** *132* (11).
78. Zhao, Y.; Schultz, N. E.; Truhlar, D. G., Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *Journal of chemical theory and computation* **2006,** *2* (2), 364-382.
79. Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian 09, revision a. 02, gaussian. *Inc., Wallingford, CT* **2009,** *200*.
80. 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* (15).
81. (a) Besler, B. H.; Merz, K. M.; Kollman, P. A., Atomic Charges Derived from Semiempirical Methods. *Journal of computational chemistry* **1990,** *11* (4), 431-439; (b) Singh, U. C.; Kollman, P. A., An Approach to Computing Electrostatic Charges for Molecules. *Journal of computational chemistry* **1984,** *5* (2), 129-145.
82. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* **1977,** *23* (3), 327-341.

83. Sagui, C.; Pedersen, L. G.; Darden, T. A., Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J Chem Phys* **2004,** *120* (1), 73-87.
84. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of chemical theory and computation* **2015,** *11* (4), 1864-1874.
85. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of chemical theory and computation* **2013,** *9* (7), 3084-3095.
86. Lifson, S., Theory of Helix-Coil Transition in Polypeptides. *J Chem Phys* **1961,** *34* (6), 1963-&.
87. Efron, B., Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion. *Ann Stat* **1986,** *14* (4), 1301-1304.
88. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophys J* **2009,** *97* (3), 853-856.
89. Karplus, M., Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J Chem Phys* **1959,** *30* (1), 11-15.
90. Hu, J. S.; Bax, A., Determination of phi and chi(1) angles in proteins from C-13-C-13 three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *Journal of the American Chemical Society* **1997,** *119* (27), 6360-6368.
91. Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L., Intrinsic backbone preferences are fully present in blocked amino acids. *P Natl Acad Sci USA* **2006,** *103* (5), 1272-1277.
92. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *Journal of Physical Chemistry B* **2000,** *104* (15), 3712-3720.
93. Swails, J. M.; York, D. M.; Roitberg, A. E., Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *Journal of chemical theory and computation* **2014,** *10* (3), 1341-1352.
94. Lipari, G.; Szabo, A., Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *Journal of the American Chemical Society* **1982,** *104* (17), 4559-4570.
95. Prompers, J. J.; Bruschweiler, R., General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *Journal of the American Chemical Society* **2002,** *124* (16), 4522-4534.
96. Gu, Y.; Li, D. W.; Bruschweiler, R., NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment. *Journal of chemical theory and computation* **2014,** *10* (6), 2599-2607.
97. (a) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M., Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. *Journal of the American Chemical Society* **2007,** *129* (15), 4724-4730; (b) Li, D. W.; Bruschweiler, R., Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *Journal of chemical theory and computation* **2011,** *7* (6), 1773-1782.
98. (a) Richardson, J. S.; Prisant, M. G.; Richardson, D. C., Crystallographic model validation: from diagnosis to healing. *Current opinion in structural biology* **2013,** *23* (5), 707-14; (b) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B., 3rd;

Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., MolProbity: More and better reference data for improved all-atom structure validation. *Protein science : a publication of the Protein Society* **2018,** *27* (1), 293-315.
99. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-2637.
100. Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J., Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009,** *25* (11), 1422-3.
101. (a) Schrauber, H.; Eisenhaber, F.; Argos, P., Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J Mol Biol* **1993,** *230* (2), 592-612; (b) Aurora, R.; Creamer, T. P.; Srinivasan, R.; Rose, G. D., Local interactions in protein folding: Lessons from the alpha-helix. *J Biol Chem* **1997,** *272* (3), 1413-1416; (c) De Maeyer, M.; Desmet, J.; Lasters, I., All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Folding & design* **1997,** *2* (1), 53-66.
102. Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C., Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *Journal of chemical theory and computation* **2006,** *2* (2), 420-433.
103. Pace, C. N.; Scholtz, J. M., A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* **1998,** *75* (1), 422-427.
104. Chothia, C.; Lesk, A. M., The Relation between the Divergence of Sequence and Structure in Proteins. *Embo J* **1986,** *5* (4), 823-826.
105. Best, R. B.; Mittal, J., Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *Journal of Physical Chemistry B* **2010,** *114* (46), 14916-14923.
106. Abascal, J. L. F.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *J Chem Phys* **2005,** *123* (23).
107. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins-Structure Function and Bioinformatics* **2010,** *78* (8), 1950-1958.
108. (a) Doig, A. J.; Baldwin, R. L., N- and C-Capping Preferences for All 20 Amino-Acids in Alpha-Helical Peptides. *Protein Science* **1995,** *4* (7), 1325-1336; (b) Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L., Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Science* **1996,** *5* (12), 2623-2637.
109. Hall, J. B.; Fushman, D., Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J Biomol Nmr* **2003,** *27* (3), 261-275.
110. Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A., Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. *Journal of the American Chemical Society* **1995,** *117* (50), 12562-12566.
111. Buck, M.; Boyd, J.; Redfield, C.; Mackenzie, D. A.; Jeenes, D. J.; Archer, D. B.; Dobson, C. M., Structural Determinants of Protein Dynamics - Analysis of N-15 Nmr Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg-White Lysozyme. *Biochemistry-Us* **1995,** *34* (12), 4041-4055.

112. Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C., Secondary structure bias in generalized born solvent models: Comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *Journal of Physical Chemistry B* **2007,** *111* (7), 1846-1857.

TABLE OF CONTENTS