JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model

Karl T. Debiec,[†,‡,§] David S. Cerutti,[∥] Lewis R. Baker,[§] Angela M. Gronenborn,[‡] David A. Case,[⊥] and Lillian T. Chong*[,§]

[†]Molecular Biophysics and Structural Biology Graduate Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

[‡]Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, United States

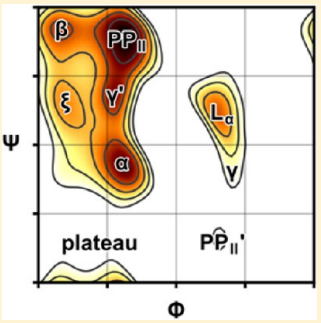
[§]Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

[∥]Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

[⊥]Department of Chemistry and Chemical Biology, Rutgers University, New Brunswick, New Jersey 08854, United States

Supporting Information

ABSTRACT: We present the AMBER ff15ipq force field for proteins, the second-generation force field developed using the Implicitly Polarized Q (IPolQ) scheme for deriving implicitly polarized atomic charges in the presence of explicit solvent. The ff15ipq force field is a complete rederivation including more than 300 unique atomic charges, 900 unique torsion terms, 60 new angle parameters, and new atomic radii for polar hydrogens. The atomic charges were derived in the context of the SPC/E_b water model, which yields more-accurate rotational diffusion of proteins and enables direct calculation of nuclear magnetic resonance (NMR) relaxation parameters from molecular dynamics simulations. The atomic radii improve the accuracy of modeling salt bridge interactions relative to contemporary fixed-charge force fields, rectifying a limitation of ff14ipq that resulted from its use of pair-specific Lennard-Jones radii. In addition, ff15ipq reproduces penta-alanine J-coupling constants exceptionally well, gives reasonable agreement with NMR relaxation rates, and maintains the expected conformational propensities of structured proteins/peptides, as well as disordered peptides—all on the microsecond (μs) time scale, which is a critical regime for drug design applications. These encouraging results demonstrate the power and robustness of our automated methods for deriving new force fields. All parameters described here and the mdgx program used to fit them are included in the AmberTools16 distribution.



1. INTRODUCTION

Akin to developments spurred by the rapid expansion of computer power around 2000, the burgeoning capacity provided by programmable graphics processing units (GPUs) has extended the utility of molecular simulations as a practical tool for assessing biophysical processes.[1,2] Notably, GPU-accelerated computing has enabled routine simulations on the microsecond (μs) time scale, a critical regime on which biological processes including protein recognition, ligand binding, and protein conformational changes occur.[3] Access to these longer time scales may reveal flaws in the simulation models that were not previously apparent, driving refinements and leading toward improved predictive power of these models.

Historically, efforts in force field development have been largely focused on selecting only a subset of the parameters in a complex model for reoptimization, e.g., reoptimizing certain torsion parameters while keeping the set of atomic charges fixed to improve the accuracy in modeling particular behaviors, while retaining what is already successful. However, such efforts are limited by the accuracy of the unoptimized parameters. Many recent force field updates have focused on refinements of torsion parameters.[4−14] However, these refinements may be compensating for deficiencies in the modeling of electrostatic and nonbonded interactions that limit the maximum attainable accuracy of the model. In addition, contemporary force fields have a tendency to borrow from a similar set of values for bond lengths, angles, and atomic radii that were fit many years ago, leading to interdependencies that may be difficult to untangle when optimizing only a portion of the parameters.

More recently, semiautomated schemes have been developed to simultaneously optimize hundreds of parameters, thereby enabling the rapid development of new force fields. In particular, the Force Balance and Implicitly Polarized Charge (IPolQ) methods[15−17] have yielded the AMBER ff15fb and ff14ipq force fields,[17] respectively. These methods rely largely on automated tools for parameter optimization, but still require some amount of manual intervention in the form of fitting set composition or user-specified settings of the fitting algorithm. The engine behind the IPolQ workflow is the mdgx module of the AMBER software package.[18] This module combines its molecular dynamics (MD) facility with linear algebra routines

Received: June 1, 2016
Published: July 11, 2016

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

for solving least-squares problems, manages extensive bookkeeping to organize parameters, provides user control over the fitting process, and interprets statistics to aid in further refinements. The mdgx module contains charge and torsion fitting routines that were built throughout the development of AMBER ff14ipq, the first complete protein force field based on the IPolQ scheme.[17]

Here, we have developed the new AMBER ff15ipq force field using the IPolQ workflow. The original motivation for the development of ff15ipq was to tackle concerns that its predecessor, ff14ipq, overestimates the stability of salt-bridge interactions—a limitation shared with many other contemporary force fields.[19] However, in contrast to recently developed variants of force fields that address such concerns,[12,20,21] ff15ipq is far from a limited adjustment of its predecessor. Rather, ff15ipq is a complete rederivation, comprising new atomic charges, a greatly expanded torsion parameter set with several new atom types to decouple distinct amino acids, and new backbone angle bending terms. In addition, whereas ff14ipq employed the TIP4P-Ew model for the solvent in the IPolQ scheme,[22] ff15ipq uses SPC/E_b, a recently developed three-point water model that yields more-accurate rotational diffusion for proteins in solution.[23] The use of a three-point water model instead of a four-point water model, leaving out a virtual site, affords a modest improvement in the speed of CPU-based simulations and a larger acceleration in computing under the AMBER GPU engine.[24] In addition, the more-accurate rotational diffusion afforded by SPC/E_b opens new avenues for validating ff15ipq through direct calculation of NMR relaxation parameters. With the aid of GPUs and the AMBER GPU engine, we have extensively validated the force field by running MD simulations of peptide and protein systems on the μs time scale, yielding over 200 μs of aggregate simulation time.

We expect ff15ipq, or a close relative, to be valuable for a long time, even as we explore more expensive alternatives by adding virtual sites to both the protein and the standard water model. In addition, we are working to apply the mdgx workflow to other classes of biopolymers such as carbohydrates and nucleic acids, and to small organic molecules. We hope that the sweeping reoptimization made possible by mdgx and tools similar to it will inspire initiatives with other force fields and create complete chemical representations with predictive power in biomolecular simulations. Of future interest will be comparisons to contemporary force fields that have been developed in the traditional manner such as AMBER ff14SB, OPLS-AA/M, and CHARMM36,[4,11,13] as well as those developed using alternative sweeping reoptimization schemes, such as AMBER ff15fb.[15]

2. THEORY

2.1. The IPolQ Method of Force Field Parameterization. The Implicitly Polarized Charge (IPolQ) method is a protocol for parametrizing fixed-charge force fields for solution-phase simulations that is comprised of two main components, implemented in the mdgx program of AmberTools.[18] The first component is a protocol for deriving nonpolarizable atomic charges that implicitly represent the energy of polarization by the presence of a solvent such as water.[16] The IPolQ charge derivation draws on approximations of dipole interactions in an external electrostatic field to arrive at the optimal nonpolarizable representation of a solute's atomic charges in the presence of a solvent such as water: precisely halfway between the charges that would reproduce the solute's electrostatic field in the gas phase and those that would reproduce the solute's electrostatic field after solvent-induced polarization.[25] This averaging comes about from the fact that the energy of a set of polarizable dipoles in an external field is identical to the energy of a set of fixed dipoles whose polarizations are halfway between the field-polarized dipoles and their gas-phase counterparts. IPolQ fits such fixed charges by applying the Restrained Electrostatic Potential (REsP) method,[26] using a pair of representations of the solute's electrostatic field corresponding to the vacuum and solution phases. While the former representation is straightforward to obtain from QM calculations, the latter is computationally unfeasible using a pure QM representation. Instead, the IPolQ method represents the polarizing Solvent Reaction Field Potential (SRFP) in its QM calculation using a field of point charges, derived from an MD simulation in which water, represented by the model with which the solute will ultimately be simulated, moves in equilibrium around the fixed solute.[16] Atomic charges are subsequently fit to reproduce the QM electrostatic potential at a set of grid points surrounding the molecule. As described previously,[16] grid points are selected within the first and second solvation shells, with the inner boundary defined by excluding points for which the energy of the Lennard-Jones interaction between the solute and a probe representing the water model exceeds a selected maximum cutoff. While, in this work, we have applied equal weights to all of the selected grid points, others have found that more consistent charges may be obtained by applying a weighting function to de-emphasize points close to or distant from the solute.[27] Such improvements will be investigated as the IPolQ method is applied to other classes of molecules beyond peptides and proteins.

The second component of the IPolQ method is an extension for the fitting of bonded parameters that accounts for the discrepancy between the desired solution-phase conformational preferences and vacuum-phase QM calculations. This is accomplished by fitting a pair of solute charge sets: one appropriate for the vacuum phase (Q_{vac}), and the other for the solution phase (Q_{solv}). In the presence of the Q_{vac} charge set, the force field's bonded parameters are fit to reproduce the relative vacuum-phase QM energies of a diverse set of solute conformations. In subsequent simulations in the solution phase, these same bonded parameters are paired with the polarized Q_{solv} charge set with the intention that the difference in the charge sets would account for the difference in solute conformational preferences between the vacuum and solution phases.[17]

2.2. Choice of Water Model for Rederivation of IPolQ Atomic Charges. In contrast to the standard REsP method of fitting atomic charges for AMBER force fields,[26] the IPolQ method explicitly considers the influence of the water model on the solute's charge distribution.[16] While the atomic charges of the ff14ipq force field were fit using the TIP4P-Ew water model,[22] we have elected to fit the charges of ff15ipq using the SPC/E_b water model.[23] This recently developed water model offers two advantages: in addition to the reduction in computational cost that is obtained by switching from a four-point water model to a three-point water model, this water model has been parametrized to yield accurate rotational diffusion of solvated proteins.

A key advantage to performing simulations with accurate rotational diffusion is the ability to directly calculate the NMR relaxation parameters ^{15}N R_1 and R_2 and ^{15}N−^{1}H heteronuclear NOE. These parameters provide information about fast

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

dynamics (picosecond (ps) or nanosecond (ns) scale) of individual backbone N−H bond vectors within a protein, potentially offering a powerful means with which to validate MD simulations.[28] In principle, these NMR relaxation parameters may be calculated directly from an MD trajectory from the autocorrelation functions of the backbone N−H vectors. In practice, however, the poor reproduction of protein rotational diffusion in MD simulations using popular water models such as TIP3P limits the utility of such calculations.[23,29] This limitation has historically been addressed using approaches such as model-free analysis that attempt to separate the global rotational diffusion of the protein from its residue-specific internal dynamics, comparing only the internal dynamics between experiment and simulation. However, these approaches require extensive fitting of models to the experimental data, and further require that the global and local dynamics occur on separable time scales, limiting their applicability to highly flexible systems such as disordered peptides and proteins. Therefore, it would be preferable for MD simulations to yield accurate rotational diffusion, such that the simulated and experimental relaxation parameters can be compared directly.

Our decision to fit the charges of ff15ipq to the SPC/E_b water model was based on preliminary tests in which we ran a series of 24 simulations of the proteins GB3, ubiquitin, and binase using two different force fields (AMBER ff99SB-ILDN and CHARMM22*)[5,12] paired with four different water models (TIP3P, TIP4P-Ew, TIP4P-D, and SPC/E_b)[22,23,30,31] (see Figure S1 in the Supporting Information). Consistent with prior published work,[23,29] TIP3P and TIP4P-Ew yielded rotational diffusion significantly faster than experiment, and SPC/E_b yielded the most accurate result.

2.3. Extensions Supporting Restrained Angle Fitting. Earlier versions of mdgx were capable of fitting angle stiffnesses alongside torsions, but recent advances proposed by Vanommeslaeghe et al. permit the calculation of both the optimal stiffness constant and equilibrium value from the same linear least-squares problem.[32] The strategy generalizes work by Hopkins and Roitberg,[33] representing the parabolic angle as the sum of two parabolic basis functions and solving for both scaling coefficients to interpolate the optimal parameters. Basis functions were chosen such that their minima lay at ±0.2 radians from the equilibria of the original angles, which had been inherited from the AMBER ff94 force field.[34] As explained by Vanommeslaeghe et al., the optimized angle's stiffness is given by the sum of coefficients solved for each basis function, and its equilibrium is given by the average of the two basis functions' minima weighted by their coefficients. Restraints on the optimized angle parameters follow from these definitions: a restraint equation setting the sum of the two coefficients $C_{i,1}$ and $C_{i,2}$ to a target value K_i, such as the stiffness of the original angle in the input force field, will harmonically penalize solutions which depart from the original stiffness value:

$$\alpha_{scl} N_i \alpha_{cpl} [C_{i,1} + C_{i,2}] = \alpha_{scl} N_i \alpha_{cpl} K_i$$

A similar restraint on the ratio of the two coefficients can be used to penalize solutions that depart from the original equilibrium value T_i, if the minima of the basis functions scaled by $C_{i,1}$ and $C_{i,2}$ are $B_{i,1}$ and $B_{i,2}$, respectively:

$$\alpha_{scl} N_i [C_{i,1}(T_i - B_{i,1}) + C_{i,2}(T_i - B_{i,2})] = 0.0$$

As explained in the previous study,[17] these restraint equations will have a more pronounced effect if the dataset contains only 10 data points rather than 1000 data points. Therefore, both sides of each restraint equation were scaled by a user-defined constant (α_{scl}) times the number of instances in which each optimizable angle appeared in the dataset N_i, analogous to the scaling constant applied to torsion restraints. The scaling constants may appear to have no effect on the solution to the equations when either these constants are present on both sides of the equation or one side of the equation is zero. However, since the least-squares fit finds an approximate solution to each equation, the scaling constants do, in fact, influence the relative importance of each restraint. In addition, because of the fact that these restraints penalize numerical deviations from the target values but angle stiffnesses and equilibria are expressed in different units by numbers of different scale, a separate scaling factor (α_{cpl}) was introduced to control the way in which restraints on the equilibria scale, relative to restraints on the stiffness. For example, an α_{cpl} of ∼57 applied to restraints on equilibria would penalize 1° deviations from the original value by the same amount as a 1 kcal/(mol rad^2) deviation from the original stiffness constant. After some experimentation, however, we found that, in our very large and heterogeneous datasets, much smaller values of α_{cpl} (0.5−1.0) result in the best fits to the data while partitioning the changes between equilibria and stiffness constants.

2.4. Addition of New Atom Types. Alongside the fitting of torsions and angles, several new atom types were added to ff15ipq in order to more accurately capture residue-specific conformational preferences. Most protein force fields use Gly and Ala as templates to develop backbone torsion parameters that are then inherited by other residues. The AMBER IPolQ force fields adopt an unconventional, concerted approach in which Φ, Ψ, Φ′, Ψ′, and all other torsions are simultaneously fit to the conformational preferences of all residues in which they appear. While the resulting backbone torsions therefore consider the conformational preferences of residues other than Gly and Ala, their overall accuracy may decrease as different residues pull the parameters in different directions. Within the context of the IPolQ fitting method, a set of backbone torsion parameters that more accurately capture the conformational preferences of different residues may be obtained by introducing new atom types, creating decoupled classes of backbone torsions, which are applied to subsets of residues.

During the development of ff14ipq, three such classes were introduced: one for Gly, one for Pro, and one for all other residues. In order to decouple the Φ and Ψ torsions of Gly, which lacks Cβ and therefore has no Φ′ and Ψ′ torsions, a unique Cα atom type was assigned.[17] Similarly, the backbone torsions of Pro were decoupled by assigning a unique atom type for the backbone N. This additional atom type not only created unique Φ, Ψ, and Ψ′ terms for Pro, but also a set of separate Ψ and Ψ′ torsions for residues preceding Pro, thereby enabling the force field to capture the unique conformational preferences of these contexts.[35] The remaining residues were further divided into three subclasses based on their Cβ types, which determine the applied Φ′ and Ψ′ torsions. The Cβ types of ff14ipq yielded four subclasses: (i) flexible positively charged residues (Arg, Lys), (ii) residues whose Cβ atoms are bonded to two heavy atoms and whose side-chains are not aromatic (Asn, Asp, Cys, Gln, Glu, Leu, Met, Ser), (iii) residues whose

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

Cβ atoms are bonded to three heavy atoms (Ile, Thr, Val), and (iv) all other residues (Ala, His, Phe, Trp, Tyr). Notably, this last subclass yielded Φ′ and Ψ′ torsions shared between Ala and the bulky aromatic residues. This unusual coupling was a consequence of the force field's lineage from ff12SB, where refitting of X_1 torsions alongside fixed Φ, Ψ, Φ′, and Ψ′ did not require a unique Cβ type for the aromatic residues, which already have unique Cγ types that yield unique X_1.[4]

In order to further improve the accuracy of residue-specific conformational preferences in ff15ipq, several new atom types were added to further decouple the backbone torsion parameters of different residues, leading to a total of five backbone classes. In order to restrict each class of backbone torsions to a single set of scaled 1−4 electrostatic terms, negatively charged (Asp, Glu) and positively charged (Arg, Lys) residues have been given unique Cα types, decoupling their Φ, Ψ, Φ′, and Ψ′ from those of the neutral residues. While the backbone N of Pro was decoupled in ff14ipq, it retained a shared Ψ′ torsion; to break this dependency, Pro has now been assigned a new Cα atom type. Finally, the coupling between Ala and the bulky aromatic residues has been removed by assigning His, Phe, Trp, and Tyr a unique Cβ type, decoupling their Φ′ and Ψ′ terms from Ala. This decoupling divides the neutral residues into four subclasses: (i) Ala, (ii) residues whose Cβ atoms are bonded to two heavy atoms (Asn, Cys, Gln, Leu, Met, Ser), (iii) residues whose Cβ atoms are bonded to three heavy atoms (Ile, Thr, Val), and (iv) bulky aromatic residues (His, Phe, Trp, Tyr). The backbone torsion classes of the 28 residue forms supported by ff15ipq are listed in Table S1 in the Supporting Information.

3. METHODS

3.1. Calculation of the Probability of Binding (P_{bound}) for Salt-Bridge Formation. To compare the accuracy of ff15ipq in modeling the stability of protein salt bridges to its predecessor ff14ipq and contemporary force fields, we simulated the association of three pairs of oppositely charged amino acid side-chain analogues: guanidinium cation/acetate anion (Arg/Asp), butylammonium cation/acetate anion (Lys/Asp), and imidazolium cation/acetate anion (His(+)/Asp). For comparison, such simulations were carried out using the polarizable force fields CHARMM Drude-2013 and AMOEBA,[36,37] in addition to other fixed-charge force fields that we had previously tested using these three model systems.[19]

Simulations with ff15ipq, ff14ipq, and AMOEBA were carried out using the AMBER 15 software package,[18] while those with CHARMM Drude-2013 were run with NAMD 2.10.0,[38,39] following a protocol analogous to that used for the previously evaluated fixed-charge force fields (full details are provided in the Supporting Information).[19] Systems were constructed to be consistent with the experimental conditions under which the association constants (K_A) of guanidinium acetate and butylammonium acetate have been measured,[40] i.e., each system consisted of 100 molecules of cation (guanidinium, butylammonium, or imidazolium), 2 molecules of acetate, and 98 chloride counterions solvated by ∼18 000 water molecules. For the fixed-charge force fields, parameters of the side-chain analogues were based on those of the complete amino acids. For the CHARMM Drude-2013 polarizable force field, parameters of guanidinium, imidazolium, and acetate were those distributed alongside the force field.[36] Since methylammonium rather than butylammonium was used as the analogue of Lys during the development of Drude-2013,[36] the butylammonium acetate system was not tested with this force field. For the AMOEBA force field, parameters of guanidinium, imidazolium, and acetate were generated using the Poltype derivation protocol (details are provided in the Supporting Information).[41] As it was with CHARMM Drude-2013, the butylammonium acetate system was not tested with AMOEBA.

For all of the simulations mentioned above, the probability that an acetate molecule was bound to one or more cation molecules was calculated by assigning each pair to either the bound or the unbound state. For each force field and pair of side-chain analogues, definitions of the unbound and bound states were based on the potential of mean force (PMF) as a function of the minimum distance between nitrogen atom(s) of the cation and the oxygen atoms of acetate. In particular, the cutoff between the bound and unbound states was defined as the point of inflection between the free energy minimum of the bound state (∼2.5−3 Å) and the free-energy maximum, which corresponds to the desolvation barrier (∼3−3.5 Å). Pairs whose minimum N−O distances were below this cutoff were assigned to the bound state, while those beyond were assigned to the unbound state. In addition to species in which a single acetate molecule was bound to a single cation molecule, forming a 1:1 complex (e.g., the guanidinium/acetate complex), species in which acetate was bound to two or more cation molecules (e.g., the 2:1 diguanidinium/acetate complex) were observed and counted separately. Standard errors were calculated using a block averaging method.[42]

3.2. Rederivation of IPolQ Atomic Charges with the SPC/E_b Water Model. The atomic charges of ff15ipq were fit using the IPolQ module of mdgx, as described previously for ff14ipq.[17] During charge fitting, each amino acid was represented by a blocked dipeptide including acetyl (Ace) and *N*-methylamide (Nme) caps; terminal forms were represented by omitting one of the blocking groups, while the disulfide form of cysteine (Cyx) was represented by a pair of dipeptides linked by a disulfide bond. To expand on the set of amino acids and protonation states that were supported by ff15ipq, atomic charges were also derived for the following: the N- and C-terminal forms of protonated aspartate (Ash) and glutamate (Glh), the C-terminal form of neutral lysine (Lyn), the terminal and nonterminal forms of deprotonated cysteine (Cym), and the noncanonical amino acid norleucine (Nle).

Each solute of interest was solvated in a cubic box of SPC/E_b water with a clearance of 10 Å between the solute and the edge of the box, and subjected to a high-temperature MD simulation at 450 K, from which were collected a set of 20 conformations. Each conformation was subsequently re-equilibrated at 298 K before being input to the IPolQ module of mdgx. This module was used to run an MD simulation with the solute fixed, during which the coordinates of surrounding solvent molecules were collected and used to generate a collection of point charges representing the solvent reaction field potential. This collection consists of an inner cloud of point charges taken directly from the coordinates of solvent molecules within 5 Å of the solute, and three outer shells of point charges fit to reproduce contributions to the solvent reaction field potential from the infinite periodic system beyond 5 Å.

A pair of QM calculations for the solute were then run at the MP2/cc-pVTZ level of theory:[43−46] one in vacuum and the other including the solvent reaction field potential. as modeled by the collection of point charges. These calculations were run using the ORCA 3.0.3 software package for each conformation of each residue,[47] requiring over 3000 density calculations. The

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

resulting densities were then input to mdgx's FitQ module, yielding a pair of charge sets: one valid for simulation under vacuum (Q_{vac}) and the other for simulation in solution (Q_{solv}).

3.3. Generation and Extension of the Angle and Torsion Fitting Dataset. The bonded parameters of ff15ipq were fit to reproduce the relative vacuum-phase QM MP2/cc-pVTZ potential energies of a set of diverse conformations of short peptides using an iterative cycle of refinement, similar to that used for its predecessor, ff14ipq.[17] This cycle involved the following steps: (i) MD simulations were carried out to generate a set of peptide conformations, (ii) these conformations were subjected to energy minimization in vacuum using the molecular mechanics (MM) energy function with Q_{vac} and the current generation of bonded parameters, (iii) QM energies of the energy-minimized conformations were calculated, (iv) the conformations and energies were used to fit an improved set of bonded parameters, and (v) steps (i) through (iv) were repeated to fit the next generation of bonded parameters. In this way, subsequent generations of the force field "learned" from the biases of their ancestors, provided those biases were captured in the QM energies of the additional conformations that resulted from step (i) of the iterative cycle.

During the development of ff14ipq, selected conformations from an initial fitting set of ∼28 000 were subjected to energy minimization with each new generation of bonded parameters to yield new conformations, accumulating a total of 65 000 structures and single-point energies.[17] The first generation of ff15ipq fitting data was created by pairing ff14ipq with generalized Born implicit solvent MD simulations[48] of amino-acid dipeptides at 450 K, followed by vacuum energy minimization of many snapshots from each simulation. While we have not tested how well ff14ipq behaves with implicit solvent, the purpose was to capture any spurious conformational preferences that might remain in the original force field. In addition, we included ∼1400 conformations of the Ace-Ala-Pro-Ala-Nme tetrapeptide, while the second generation added numerous tripeptides containing Gly, and conformations of the disulfide-bridged Cys·Cys system (among the largest of all the systems used in QM single-point-energy calculations). These refinements added ∼15 000 new conformations to the ff15ipq fitting set.

The next three generations of refinement were designed to cover sampling of the multiple classes of backbone parameters applied to different residues, as described in section 2.4. In order to ensure sampling of diverse backbone conformations, conformations were generated by progressively restraining Φ and Ψ at 20° intervals, using a 16 kcal/mol·rad^2 harmonic restraint over the course of the MD simulation, yielding 324 conformations of each. Since the unique backbone nitrogen type of Pro creates unique Ψ and Ψ′ terms for preceding residues, the third generation of conformations consisted of 51 Pro-containing tripeptides for which the non-Pro residue's Φ and Ψ were restrained. At this point in development, it was decided to branch the positively- and negatively charged residues into unique backbone classes, and as such the fourth and fifth generations of refinement consisted of 57 tripeptides containing the charged residues Asp, Glu, Cym (deprotonated Cys), Arg, Lys, and Hip (doubly protonated His), in which Φ and Ψ of the charged residues were restrained. In order to cover the unique backbone parameters applied to the terminal forms of each residue, additional conformations were added for a set of 78 terminal NXaa-Nme and Ace-CXaa monopeptides. For these terminal systems, scans of either the unique Ψ of the N-terminal forms or the unique Φ of the C-terminal forms were performed at 2° intervals, yielding 180 conformations of each. Since the unique backbone nitrogen types of the N-termini and Pro in tandem yield an additional set of Ψ terms for NXaa-Pro, scans of Ψ were run for an additional set of NXaa-Pro-Nme dipeptides. Finally, in order to cover the unique backbone Ψ and Ψ′ terms of the amide blocking group (Nhe), scans of Φ and Ψ were run for 17 Ace-Xaa-Nhe dipeptides, yielding 324 conformations of each. During these three generations of refinement, ∼60 000 conformations were added to the ff15ipq fitting set.

After the fifth generation of refinement, support for the fitting of angle equilibria and force constants alongside torsions was implemented in mdgx, and subsequent generations emphasized comprehensive sampling of backbone angles. The sixth, seventh, and eighth generations of refinement consisted of perturbations of the angles around N, Cα, and C. Starting from an initial conformation, a selected angle of interest was subjected to a random perturbation within a range of ±20° of its original equilibrium value (as inherited from the ff94 force field and retained in contemporaries such as ff14SB). Target values for the other angles around the same central atom were then chosen by taking their initial values and adjusting them such that the total sum of angles around the central atom was appropriate for the known geometry; target sums of 360° for planar geometry around N and C and 660° for tetrahedral geometry around Cα were used. During subsequent MM minimization, the target values for these angles were restrained using 256 kcal/mol rad^2 harmonic restraints. During the eighth and final generation of refinement, angle perturbations were resampled in the context of new scans of Φ and Ψ backbone torsions at 10° intervals for each Ace-Xaa-Nme dipeptide, yielding 1296 conformations of each, alongside additional sampling of terminal monopeptides. During these three generations, ∼125 000 conformations were added, yielding a final fitting set of ff15ipq consisting of >250 000 single-point QM energies, which is over four times larger than that used for ff14ipq.

3.4. Fitting of Torsion and Angle Terms. As done previously for ff14ipq, the torsion parameters of ff15ipq were fit using a linear least-squares fit implemented in the Param module of mdgx;[17] extensions to the module for angle fitting are described in section 2.2. This module selects a set of torsional barrier heights, angle equilibria, and angle stiffnesses that best reproduce the relative conformational energies of the systems included in the fitting set. During the fitting process, the Fourier series lengths and phase angles of the torsional terms were not optimized, and phase angles were set to either 0° or 180° to enable the development of parameters that are transferable to alternative chiralities. All backbone Φ, Ψ, Φ′, Ψ′, and side-chain X torsions of nonterminal forms of the amino acid residues were allocated four terms in their Fourier series. Torsions unique to the terminal forms of residues and residues preceding Pro were restricted to only three terms since these terms were less exhaustively sampled in the fitting set. This restriction was applied to limit the risk of overfitting. While all torsion parameters of residues in the fitting set were fit, only angles in which the central atom was the N, Cα, or C of a nonterminal residue were fit. For Pro, only the angles around Cα were fit, since the unique backbone N type of Pro introduces a large number of parameters that are dependent on the preceding residue. In order to avoid overfitting torsional barrier heights, torsions were restrained toward 0° with a force

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

Table 1. Peptide and Protein Validation Systems

system	sequence/PDB	residues	temperature (K)	duration (μs)
Ala_5	$^{+}$AAAAA$^{\varnothing}$	5	298	6
K19	Ace-GGG-$(KAAAA)_3$-K-Nhe (from ref 51)	19	275, 285, ..., 315, 325	4
$(AAQAA)_3$	Ace-$(AAQAA)_3$-Nhe (from ref 52)	15	280, 290, ..., 320, 330	4
GB1 hairpin	$^{+}$GEWTYDDATKTFTVTE^{-} (from ref 53)	16	275, 285, ..., 315, 325	4
chignolin	$^{+}$GYDPETGTWG^{-}, 1UAO (from ref 54)	10	298	4
Cln025	$^{+}$YYDPETGTWY^{-}, 2RVD (from ref 55)	10	280, 290, ..., 360, 370	4
Trp-cage	1L2Y (from ref 56)	20	275, 285, ..., 315, 325	4
binase	1BUJ (from ref 57)	109	298	10
BPTI	5PTI (from ref 58)	58	298	10
GB3	1P7E (from ref 59)	56	298	10
lysozyme	4LZT (from ref 60)	129	300	2
ubiquitin	1UBQ (from ref 61)	76	298	10
villin headpiece[a]	2F4K (from ref 62)	35	303	10
P53[b]	1YCR (from ref 63)	13	298	10
P53/MDM2[b]	1YCR (from ref 63)	13/85	298	10
S-peptide[c]	1RNU (from ref 64)	22	298	10
S-peptide/S-protein[c]	1RNU (from ref 64)	22/104	298	10

[a]HP35 double-norleucine mutant mutant (Lys24Nle, Asn27His, and Lys29Nle). [b]The p53 peptide used contained residues 17−29 of the full-length protein and included an N-terminal acetyl (Ace) and C-terminal amide (Nhe) blocking group. MDM2 included residues 25−109 of the full-length protein, omitting a mobile N-terminal region unresolved in the crystal structure. The N- and C-termini of MDM2 were blocked with acetyl (Ace) and *N*-methylamide (Nme) blocking groups, respectively. [c]In order to accurately match the amino acid sequences used in NMR experiments,[106,107] residues not resolved in the crystal structure were built using Avogadro.[108] Residues STSAA were appended to the C-terminus of the S-peptide, and SSS to the N-terminus of the S-protein. In addition, GA residues were appended to the N-terminus of the S-peptide, representing a cloning artifact present in the NMR experiments. These residues were not restrained during equilibration of the system. Given that the NMR experiments on the S-peptide/S-protein complex were conducted at pH 3.7, which is close to the average pK_a values of Asp (∼3.7) and Glu (∼4.1), we elected to run our simulations with negatively charged Asp and neutral Glu.

constant of 2×10^{-4} kcal/mol. Similarly, angles were restrained to their original values, inherited from ff94, with the equilibria and stiffness force constants set to 5×10^{-5} kcal/mol and 2×10^{-4} kcal/mol, respectively.

3.5. Umbrella Sampling of Tetrapeptides. To characterize the backbone conformational preferences of ff15ipq in explicit SPC/E_b water, we carried out umbrella sampling simulations of blocked tetrapeptides Ace-Ala-Xaa-Ala-Nme, calculating the potential of mean force as a function of the backbone Φ and Ψ torsions of the central residue Xaa. In order to identify differences in the conformational preferences between ff15ipq, its predecessor ff14ipq, and contemporary force fields, simulations of Ace-Ala-Ala-Ala-Nme were carried out using the AMBER force fields ff15ipq, ff14ipq,[17] and ff14SB;[4] the OPLS force field OPLS-AA/M;[13] and the CHARMM force fields CHARMM36 and Drude-2013.[11,36] Analogous simulations were carried out to compare the conformational preferences of other central amino acid residues using the AMBER ff15ipq, ff14ipq, ff14SB, and CHARMM36 force fields. The backbone Φ and Ψ torsions of the central residue were restrained in a series of 1296 windows spaced at 10° intervals, using a harmonic penalty function with a force constant of 8 kcal/mol rad^2. Each window was seeded from a continuous, incrementally restrained simulation, and sampled for 2.0 ns, following a 0.2 ns equilibration. From each set of 1296 windows were reconstructed the unbiased potentials of mean force using the weighted histogram analysis method (WHAM).[49,50]

3.6. Simulations of Benchmark Systems. To validate ff15ipq as a general force field for peptides and proteins, extensive MD simulations on the μs time scale were carried out for a variety of benchmark systems consisting of both structured and disordered peptides and proteins. For each system, the amino-acid sequence or PDB code, sources of initial coordinates, and temperatures maintained throughout the simulations are listed in Table 1. Further details of the benchmark systems are provided below.

3.6.1. Structured Peptides and Proteins. As done for ff14ipq,[17] we validated ff15ipq by simulating penta-alanine (Ala_5), the α-helical K19 peptide, the GB1 β-hairpin from the C-terminal fragment of Protein G, the designed β-hairpin chignolin, Trp-cage, GB3, and lysozyme. We also carried out simulations of the α-helical $(AAQAA)_3$ peptide, the Cln025 mutant of the chignolin β-hairpin, the double-norleucine variant of the villin headpiece subdomain, bovine pancreatic trypsin inhibitor (BPTI), ubiquitin, and binase.

3.6.2. Disordered Peptides. In order to evaluate the ability of ff15ipq to model disordered proteins, we simulated two classic systems for studying the binding processes of disordered peptides that fold only upon binding their partner proteins: (a) the N-terminal p53 peptide and MDM2 oncoprotein, and (b) the S-peptide and S-protein cleavage products of the RNase A protein. Both of these peptides fold into α-helical conformations only upon binding their partner proteins. Simulations were performed with these peptides both in isolation and in complex with their protein binding partners.

Simulations of benchmark systems were carried out using the GPU implementation of the pmemd module in the AMBER 15 software package.[24,65] Each system was solvated in a truncated octahedral box of SPC/E_b explicit water with a 12 Å buffer for the disordered peptide/protein systems and 10 Å buffer for all other systems. Prior to production simulation, each system was subjected to energy minimization, followed by a three-stage equilibration. In the first stage, a 20 ps simulation of the energy-minimized system was carried out at constant temperature while restraining the solute heavy atoms to their initial positions

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

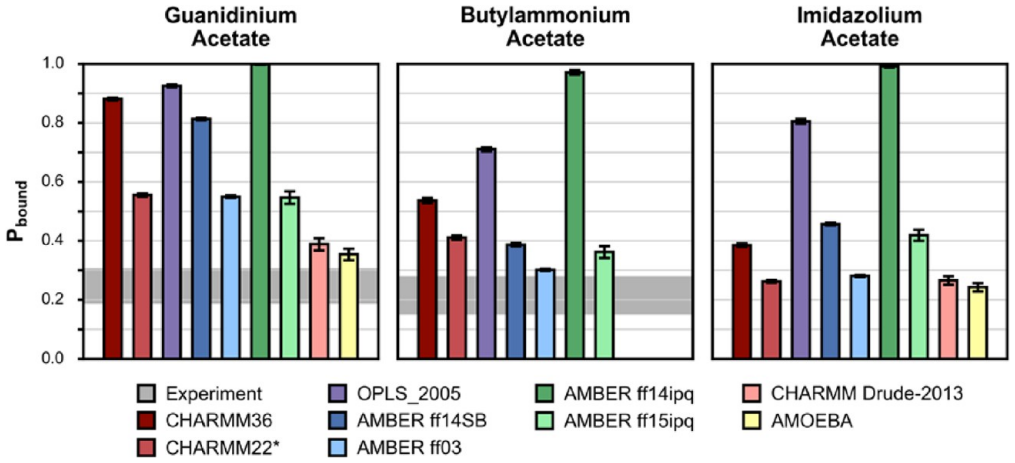


Figure 1. Probability of binding (P_{bound}) between acetate and one or more molecules of three cationic side-chain analogues using seven fixed-charge and two polarizable biomolecular force fields, each paired with either the water model with which it was derived or that with which it is most-commonly used. The P_{bound} values corresponding to the experimentally determined K_A values of guanidinium acetate and butylammonium acetate are depicted as horizontal gray bars;[40,76] no experimental value is available for the imidazolium acetate system. Error bars represent 95% confidence intervals calculated using a block averaging method.[42] Results for the CHARMM36, CHARMM22*, OPLS_2005, AMBER ff14SB, and AMBER ff03 force fields are taken from previous simulation studies.[19]

using a harmonic potential with a force constant of 1 kcal/mol Å^2. In the second stage, a 1 ns simulation was carried out at constant pressure with the same harmonic position restraints. Finally, an additional 1 ns unrestrained simulation was carried out at constant temperature and pressure. Temperatures were maintained at selected values (between 270 K and 370 K), using a Langevin thermostat (frictional constant of 1 ps^{-1}), while pressure was maintained at 1 atm using a Monte Carlo barostat (200 fs between attempts to change the system volume).[66] van der Waals and short-range electrostatic interactions were truncated at 10 Å; long-range electrostatic interactions were calculated using the particle mesh Ewald method.[67] To enable at least a 2 fs time step, bonds to hydrogen were constrained to their equilibrium values using the SHAKE and SETTLE algorithms.[68,69] For the K19, $(AAQAA)_3$, GB1 hairpin, chignolin, and Cln025 systems, hydrogen mass repartitioning was used to enable the use of longer timesteps.[70] In particular, the masses of solute hydrogen atoms were increased by a factor of 3, and that of their attached heavy atoms decreased by a corresponding amount such that the total mass remained constant; the masses of water molecules were not repartitioned. This mass repartitioning scheme enables a 4 fs time step for simulations at ≤300 K; for simulations at >300 K, shorter timesteps were used and set to be equal to 1200 K fs, divided by the set temperature. Conformations were saved every ps for analysis by the AmberTools cpptraj program.[71] Diagnostics included DSSP,[72] rotational diffusion,[29] and NMR relaxation calculated by the iRED method.[73]

4. RESULTS

4.1. Strengths of Protein Salt Bridges. To evaluate the accuracy of ff15ipq in modeling the strengths of protein salt bridges, we simulated the association of three pairs of oppositely charged amino acid side-chain analogues: guanidinium cation/acetate anion (Arg/Asp), butylammonium cation/acetate anion (Lys/Asp), and imidazolium cation/acetate anion (His(+)/Asp). For each salt bridge, the resulting probability of an anion binding to one or more cation molecules (P_{bound}) was compared to experiment (if available), as well as those of six other fixed-charge force fields, including ff14ipq, ff14SB, ff03, CHARMM22*, CHARMM36, and OPLS_2005, and two polarizable force fields (CHARMM Drude-2013 and AMOEBA).

The original motivation for the development of ff15ipq was to correct for the overstabilization of protein salt bridges by its predecessor, ff14ipq. During the development of ff14ipq, the Lennard-Jones radii of several polar heavy atoms were refit to reproduce the experimental solvation free energies of side-chain analogues.[16] Although the resulting set of radii were initially intended to be applied globally, several of the larger radii resulted in increased 1−4 repulsion during torsion fitting, which made the torsion parameters more difficult to fit. To overcome this difficulty, mixed Lennard-Jones combining rules (called LJEDIT within AMBER software or NBFIX within CHARMM) were applied to these polar groups, assigning different radii for their solute−solvent interactions from those used for their solute−solute interactions. For example, for the carboxylate oxygen atoms of the side-chains of Asp and Glu larger Lennard-Jones radii for interactions with water were used than for interactions with solute atoms.[17] An undesirable effect of this strategy, however, was the overstabilization of salt bridges—to the point that in our simulations each acetate molecule was bound to three or more cation molecules (e.g., guanidinium) for most of the simulation (see Figure 1, as well as Figure S2 in the Supporting Information). Essentially, the larger radii used for solute−solvent interactions forced the carboxylate group out of solution and into interactions with available solute atoms.

For ff15ipq, we addressed the problem of overstabilized salt bridges by discarding the mixed Lennard-Jones radii of ff14ipq and instead applied empirical corrections to the radii of polar hydrogen atoms bonded to nitrogen (atom type "H") in both the protein backbone and side-chains (note that the original σ value of 1.07 Å for this atom type may equivalently be expressed as an R* of 0.6000, and the details of its fitting appear to have been lost to history).[34,74] These corrections were determined from simulations of the three oppositely charged side-chain analogue systems with H σ ranging from the ff94 value of 1.07 Å up to 1.5 Å, calculating the probability of salt bridge formation (P_{bound}), and comparing this probability to

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

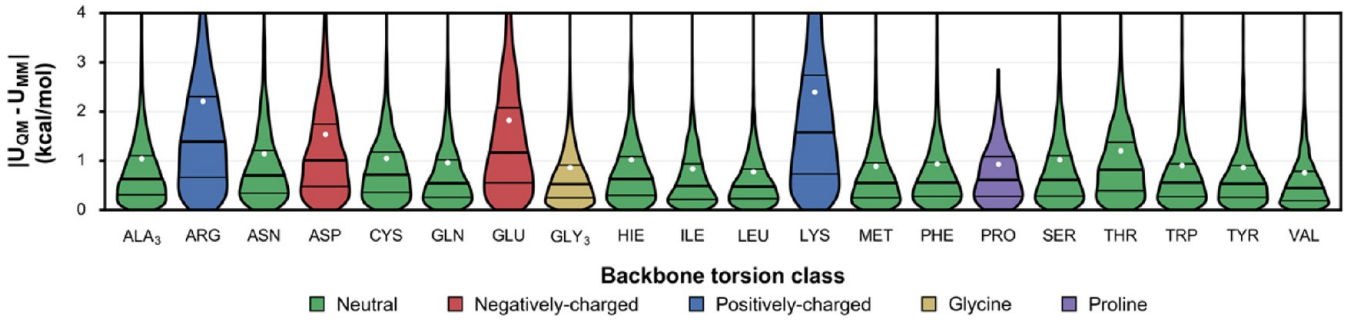


Figure 2. Distributions of residuals of relative molecular mechanical energies, with respect to their QM target potential energies, for 18 Ace-Xaa-Nme dipeptides and the Ace-Ala-Ala-Ala-Nme and Ace-Gly-Gly-Gly-Nme tetrapeptides. The 25th, 50th, and 75th percentiles are represented by horizontal lines, and root-mean-square values are represented by white circles. Each dataset is colored based on its corresponding backbone torsion class.

that from experiments. Based on the results (Figure S2), we selected a σ value of 1.3 Å for nitrogen-attached hydrogens in both the protein backbone and side-chains. For guanidinium acetate, we found that a further increase in σ to 1.5 Å for the side chain of Arg was necessary to achieve satisfactory agreement with the experimental value of this system. All other Lennard-Jones radii retained their original ff94 values.

As shown in Figure 1, ff15ipq yields P_{bound} values that are among the most reasonable, relative to the other fixed-charge force fields that were tested (ff03, ff14SB, ff14ipq, OPLS_2005, CHARMM22*, and CHARMM36). For guanidinium acetate, our results with ff15ipq are roughly consistent with CHARMM22* and ff03, which we had previously found to provide the most accurate modeling of this system (note that the corresponding figure of our previous work omitted complexes including multiple cation molecules bound to a single anion molecule and therefore under-represented the extent to which the force fields overstabilize salt bridges).[19] For butylammonium acetate, ff15ipq yields a P_{bound} value that is slightly higher than that of AMBER ff03, but similar to those of CHARMM22* and ff14SB. For imidazolium acetate, ff15ipq yields a P_{bound} that is higher than those of CHARMM22* and ff03, but similar to that of ff14SB. Of particular note is that CHARMM22* was also parametrized to reproduce the experimental association of guanidinium acetate, but via adjustments to the atomic charges of only the side-chains of Arg, Asp, and Glu;[12] as a result, these adjusted parameters are inconsistent with the rest of the force field, whose charges had been fit years earlier, using a different method.[75] Along the same lines, a recently developed variant of the AMBER ff99SB-ILDN force field has involved the application of mixed Lennard-Jones combining rules exclusively to interactions between the side chains of Arg, Asp, and Glu.[20,21] In contrast to the posthoc adjustments of these two other force fields, our approach involves first adjusting the Lennard-Jones radii, followed by refitting of atomic charges and bonded parameters. This approach—which has been an onerous one in the past—has been significantly streamlined by the mdgx software.

We note that all of the fixed-charge force fields are outperformed by the more expensive, polarizable CHARMM Drude-2013 and AMOEBA force fields. While it is likely that much of their superior performance results from the more complex charge model for the solutes, it is also possible that the solute−solvent interactions—which compete with solute−solute interactions—are more accurately represented by the use of polarizable water models. In particular, fixed-charge water models similar to the SPC/E_b model used here have recently been found to generally underestimate the strength of solute−solvent interactions,[31] and it is possible that this limitation of the water models restricts the accuracy with which the solute models may represent salt bridges.

4.2. Optimization of Torsion and Angle Parameters. A key metric for assessing the accuracy of the torsion and angle parameters of ff15ipq was the ability to reproduce the target QM potential energy surface. Figure 2 shows the distribution and root-mean-square error (RMSE) of ff15ipq energies, with respect to their target QM potential energies for the 20 canonical amino acids. The RMSE values for all neutral residues are <1.3 kcal/mol, while those of the negatively charged residues Asp and Glu are <1.9 kcal/mol, and those of the positively charged residues Arg and Lys are <2.4 kcal/mol. As shown in the Supporting Information (Figure S3), the neutral forms Asp, Gln, and Lys (Ash, Glh, and Lyn, respectively) have RMSE values that are consistent with the other neutral residues, suggesting that the increased RMSE values of Lys and Arg, relative to uncharged residues, are related to their net charge, rather than to their additional flexible X torsions. Optimization of the backbone angle parameters introduced a 5%−15% improvement in RMSE and enabled expansion of the fitting set by more than 4-fold without sacrificing the ability to reproduce those parts of the QM potential energy surface represented in the original dataset (see the section entitled "Timeline of Development" in the Supporting Information).

4.3. Conformational Preferences of Individual Residues and Very Short Peptides. To assess the accuracy of ff15ipq in modeling the backbone conformational preferences of proteins within computationally tractable systems, we carried out a series of simulations of short peptides that may be affordably simulated to convergence. Initially, we focused on simulations of the Ace-Ala-Ala-Ala-Nme tetrapeptide, for which we calculated the PMF as a function of the backbone Φ and Ψ torsions of the central residue. Figure 3 shows the results for ff15ipq, its predecessor ff14ipq, and several contemporary force fields. Relative to ff14ipq, ff15ipq has larger free energy barriers (by ∼1 kcal mol^{-1}) between the α well ($\Phi \approx -70°$, $\Psi \approx -20°$) and γ' well ($\Phi \approx -80°$, $\Psi \approx 60°$) and between the β well ($\Phi \approx -150°$, $\Psi \approx 150°$) and PPII well ($\Phi \approx -70°$, $\Psi \approx 140°$). In addition, ff15ipq has a more clearly defined ξ well ($\Phi \approx -140°$, $\Psi \approx 50°$). On the left half of the Ramachandran plot, the depth of the Lα well ($\Phi \approx 60°$, $\Psi \approx 40°$) has decreased slightly, and that of the γ well ($\Phi \approx 70°$, $\Psi \approx -40°$) has decreased by ∼1 kcal/mol, while the PII′ well ($\Phi \approx 60°$, $\Psi \approx -130°$) has been retained. Relative to the ff14SB and CHARMM36 force fields, ff15ipq shows similar α and PPII well depths, although ff14SB

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

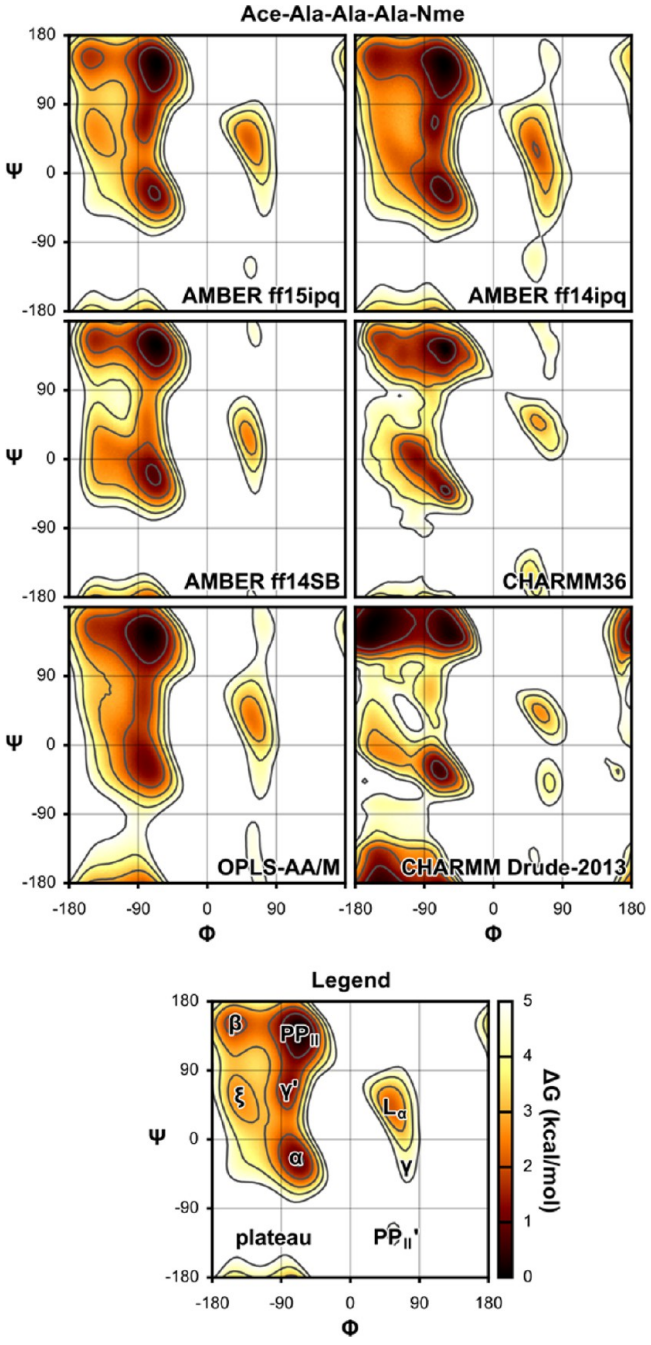


Figure 3. Potentials of mean force for the central residue of blocked alanine tetrapeptides, as a function of backbone Φ and Ψ torsions of the central residue using five fixed-charge force fields and one polarizable force field. Each force field was paired with either the water model with which it was derived or that with which it is most commonly used.

and CHARMM36 do not exhibit γ' or ξ wells and the precise positions of the various wells differ between the force fields. Larger differences are observed relative to the OPLS-AA/M and polarizable CHARMM Drude-2013 force fields, which have shallower and deeper β wells, respectively.

Next, we extended our validation of ff15ipq by examining residue-specific backbone conformational preferences. In particular, we carried out simulations of Ace-Ala-Xaa-Ala-Nme tetrapeptides containing each of the 20 canonical residues at the central position, including the 25 protonation states of these residues that are supported by the force field. For comparison, analogous simulations were carried out using the ff14ipq, ff14SB, and CHARMM36 force fields. The resulting Φ/Ψ backbone torsional preferences of the central residues were then compared to those of Ala-Xaa-Ala obtained from the Neighbor-Dependent Ramachandran Distribution (NDRD) dataset, derived from conformations observed in the loop regions of proteins (non-α-helix/β-sheet secondary structures).[77] The NDRD dataset is drawn from a collection of ∼3000 high-resolution crystal structures in the Protein Data Bank,[78] and accounts for the influence of preceding and following residues on the Φ/Ψ backbone torsional preferences of the central residue. Given the considerable differences in the contexts of our simulations and the NDRD experimental data set, i.e., solution vs crystal environment, we focused solely on qualitative differences between the simulated and experimental conformational preferences of each peptide. In particular, we compared the conformational preferences of peptides containing a nonalanine central residue, relative to that of the reference Ace-Ala-Ala-Ala-Nme peptide.

Generally, both ff15ipq and ff14ipq show greater variation between amino acids than ff14SB and CHARMM36, which apply the same backbone torsions to all residues (Figure S4 in the Supporting Information). Several differences between ff15ipq and ff14ipq are apparent. For the neutral residues whose Cβ atoms are bound to two heavy atoms, the clearest difference is the decreased favorability of the $-180° < \Psi < -90°$ region for Asn, Gln, Leu, and Met; ff15ipq is more consistent with NDRD distributions in which such conformations are rare, because of the broader sampling of such uncommon backbone conformations in the ff15ipq fitting set. An exception is Ser, which retains this region and exhibits overall broader sampling, in contrast to the NDRD distribution, in which conformations are restricted largely to the canonical wells. For the neutral residues whose Cβ atoms are bound to three heavy atoms (Ile, Thr, and Val), the NDRD dataset shows increased conformational preferences in the β region and in the region adjacent to the α well, centered at $\Phi \approx -120°$, $\Psi \approx -60°$. These preferences are captured by both ff14ipq and ff15ipq, but the lower region is erroneously disfavored by both ff14SB and CHARMM36. While ff15ipq is improved relative to ff14ipq by disfavoring the Lα well of Thr, the near-absence of sampling of this well in the NDRD distribution suggests that it may still be too favorable, relative to the experiment. Differences between ff15ipq and ff14ipq for the bulky aromatic residues are less pronounced; the conformational preferences of these residues may be more dependent on sterics as modeled by the Lennard-Jones parameters, which have not been changed in ff15ipq from those of ff14ipq. The greatest differences between ff15ipq and ff14ipq are observed for the negatively charged residues Asp and Glu, which have been granted their own Φ, Ψ, Φ′, and Φ′ torsions in ff15ipq. These residues largely restrict sampling to the of PPII, γ', and α wells, lacking clearly defined β wells and any wells on the right side of the Ramachandran plot. The differences for the positively charged residues Arg and Lys are much smaller, likely because these residues were already assigned unique Φ′ and Ψ′ torsions in ff14ipq.

To complement the above qualitative comparisons, we obtained quantitative measures of the accuracy of ff15ipq's backbone conformational preferences by calculating J-coupling constants for the Ala_5 peptide and comparing these values to experiment. This peptide was the focus of a study by Best et

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

Table 2. Ala_5 J-Coupling Constants

		Simulation				
J-coupling	residue	orig	DFT-1	DFT-2	KLL	experiment
$^1J_{N,C\alpha}$	2	11.38	11.38	11.38	11.38	11.36
$^1J_{N,C\alpha}$	3	11.06	11.06	11.06	11.06	11.26
$^2J_{N,C\alpha}$	2	8.64	8.64	8.64	8.64	9.20
$^2J_{N,C\alpha}$	3	8.50	8.50	8.50	8.50	8.55
$^3J_{C,C}$	2	0.67	0.48	0.57	0.67	0.19
$^3J_{H\alpha,C}$	2	1.57	1.31	1.47	1.38	1.85
$^3J_{H\alpha,C}$	3	1.83	1.60	1.77	1.67	1.86
$^3J_{HN,C}$	2	1.26	1.27	0.88	1.46	1.10
$^3J_{HN,C}$	3	1.19	1.20	0.88	1.37	1.15
$^3J_{HN,C\beta}$	2	2.10	4.06	3.24	2.17	2.30
$^3J_{HN,C\beta}$	3	1.99	3.79	3.02	2.04	2.24
$^3J_{HN,H\alpha}$	2	5.35	4.78	5.50	4.92	5.59
$^3J_{HN,H\alpha}$	3	5.73	5.29	5.92	5.36	5.74
$^3J_{HN,C\alpha}$	2	0.63	0.63	0.63	0.63	0.67
$^3J_{HN,C\alpha}$	3	0.62	0.62	0.62	0.62	0.68
χ^2 [a]		0.53 ± 0.02	2.91 ± 0.06	1.08 ± 0.03	0.67 ± 0.02	

[a]Uncertainties on χ^2 values represent one standard error of the mean, calculated using a block averaging method.[42] Uncertainties on individual J-coupling constants are omitted for the sake of clarity.

al.[79] in which multiple force fields were compared, in terms of their ability to reproduce experimental J-coupling constants using the Karplus equation and three different sets of Karplus coefficients: the original coefficients, as used by Graf et al.,[80−83] and two sets of DFT-based coefficients (DFT-1 and DFT-2) by Case et al.[84] In this study, a suggested criterion for a high-quality force field is that the χ^2 value between calculated and experimental J-coupling constants should be ≤2.25 for all three sets of Karplus coefficients. Three useful points of reference are the recently developed ff14SB, CHARMM36, and ff03w force fields, which were empirically corrected to improve reproduction of experimental Ala_5 J-coupling constants.[4,9,11] The ff14SB force field yielded χ^2 values of 0.9 and 1.2 with the original and DFT-2 coefficients, respectively, but a higher χ^2 of 2.7 with the DFT-1 coefficients; CHARMM36 and ff03w were tested only with the DFT-2 coefficients, yielding χ^2 values of 1.16 and 0.9, respectively.

As shown in Table 2, ff15ipq yields values of χ^2 = 0.53 with the original coefficients, χ^2 = 1.08 with the DFT-2 coefficients, and χ^2 = 0.67 with an additional set of Karplus coefficients from Lindorff-Larsen et al.[85] However, similar to ff14SB, we obtained a higher χ^2 value of 2.91 with the DFT-1 coefficients; in our case, the higher χ^2 value is driven primarily by a single outlier that deviates greatly from the experiment, $^3J_{HNC\beta}$. Based on results from preliminary versions of ff15ipq (see the "Timeline of Development" section in the Supporting Information), it appears that lower χ^2 values with the original Karplus coefficients may come at the expense of higher χ^2 values with the DFT-1 coefficients. Notably, ff15ipq—which employs a general parametrization to reproduce QM potential energies—performs at least as well as ff14SB, CHARMM36, and ff03w, which have been parametrized specifically to reproduce Ala_5 J-coupling constants.[4,11] All four force fields yield results that are much improved, relative to force fields developed only a few years ago.[79]

Note that the calculated J-couplings are dependent on the ratios of various backbone conformations, between which transitions may be relatively rare. In our studies of Ala_5, we found that 1.5 μs of aggregate simulation time, which we had previously used in our validation of ff14ipq,[17] did not yield sufficiently precise calculations of the J-couplings. Although the J-couplings may appear to be converged based on their relatively small statistical variances (evaluated using block averaging), these variances may be misleading. For example, we observed what appeared to be small, but statistically significant differences in the J-couplings between simulations run with and without hydrogen mass repartitioning after 1.5 μs of simulation, but these differences ultimately disappeared after 6 μs (see the "Timeline of Development" section in the Supporting Information). The extensive sampling needed to obtain converged J-couplings illustrates the challenge of mapping the conformations of just a few residues using brute-force MD simulation.

4.4. α-Helices: K19 and $(AAQAA)_3$ Peptides. To assess the propensity of ff15ipq to form α-helices, we studied the temperature-dependent behavior of two model α-helical peptides: K19 and $(AAQAA)_3$.[51,52] Both peptides are variants of the motif $(Ala\text{-}Ala\text{-}Xaa\text{-}Ala\text{-}Ala)_n$, in which Xaa is Lys in K19 and Gln in $(AAQAA)_3$; their sequences are listed in Table 1. For each peptide, we carried out six 4-μs simulations at different temperatures and monitored the formation of various types of secondary structure. As shown in Figure 4, both peptides undergo multiple folding and unfolding events, although our simulations are not sufficiently long to obtain converged estimates of the probability of adopting α-helical conformations. Qualitatively, K19 adopts α-helical conformations for a greater proportion of the simulation than $(AAQAA)_3$, which is consistent with the experimental observation that K19 and $(AAQAA)_3$ are ∼40% and ∼20% α-helical, respectively, at 300 K.[51,52] Both peptides transiently form β-sheet contacts, which do not appear to be stable for more than 100 ns, indicating that ff15ipq correctly identifies the favored secondary structures of these peptides.

While the two peptides differ in length, the observed difference in α-helical stability is likely due to parameters of Gln and Lys residues at the central positions of the peptides. In our umbrella sampling simulations of tetrapeptides, which are too small to form an α-helix (Figure S4), we observed a broader,

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

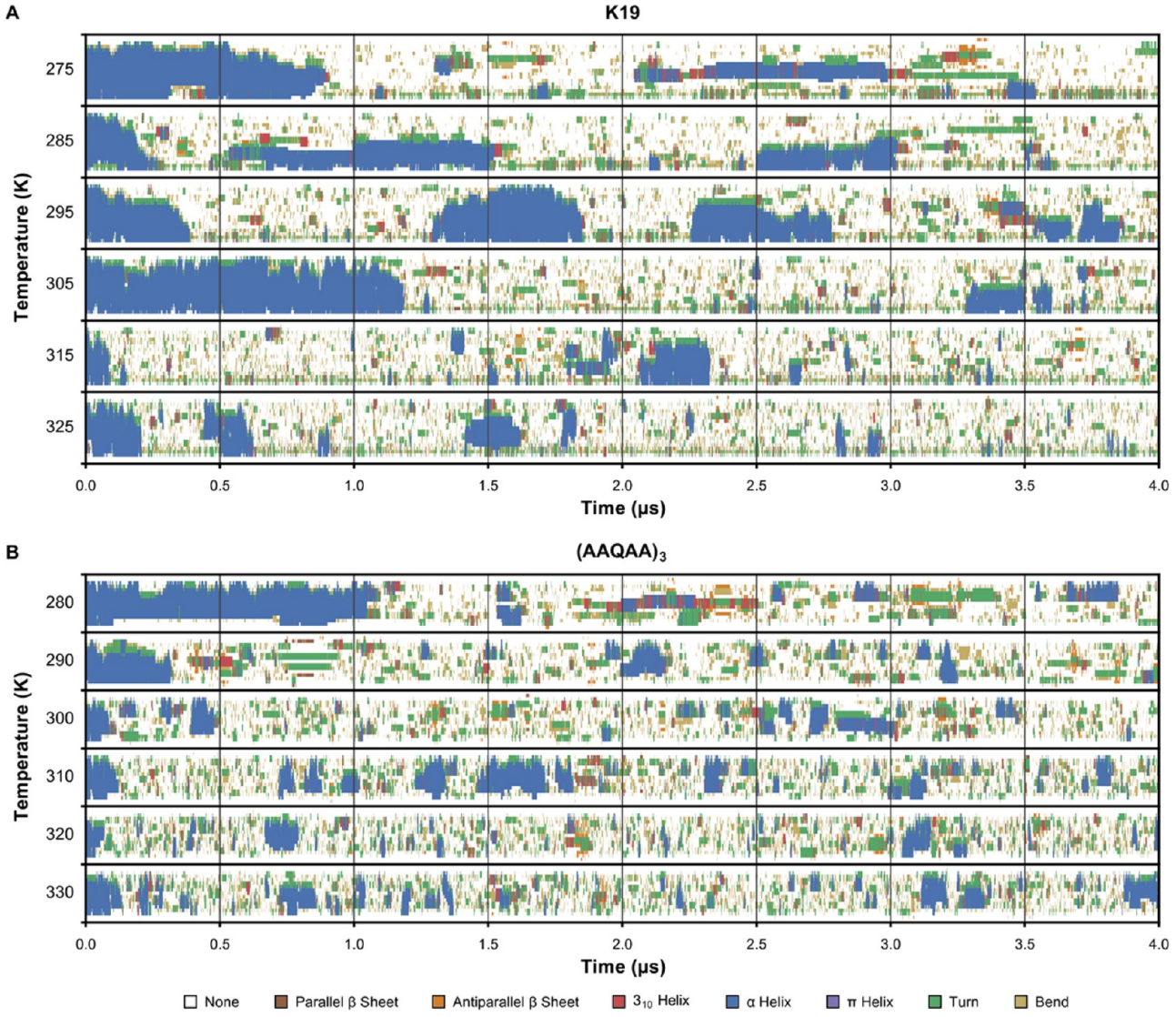


Figure 4. Secondary structure of model α-helical peptides (A) K19 and (B) $(AAQAA)_3$ at various simulated temperatures over the course of 4-μs simulations.

deeper α well for Lys than for Gln, suggesting that the observed difference in α-helical stability between the two peptides has already been "built-in" to ff15ipq at the residue level. In addition, the two residues have different backbone charges: Gln shares its N, H, C, and O charges with the other neutral residues, and Lys shares its charges with the positively charged residues. While the backbone H charges for neutral and positively charged residues are similar, the backbone O of the positively charged residues is ~0.05 e more negative than that of the neutral residues, which may result in more stable hydrogen bonding.

4.5. β-Sheets: GB1 Hairpin, Chignolin, and Cln025 Peptides. In order to assess the stability of β-sheet structures in ff15ipq, we simulated three model β-hairpin systems: the GB1 hairpin, the designed peptide chignolin, and its hyperstable variant Cln025.[53−55] We simulated the GB1 hairpin at 6 temperatures, ranging from 275 to 325 K, and we simulated Cln025 at 10 temperatures, ranging from 280 K to 370 K; we simulated chignolin only at 298 K. Figure 5 shows the secondary structures observed during 4-μs simulations of these systems. As with our simulations of the α-helical peptides, our β-hairpin simulations are not sufficiently long to precisely quantify secondary structure stability, although qualitative trends may be identified. As shown in Figure 5A, the GB1 hairpin is metastable over the tested temperature range of 275−325 K, and in two of our simulations unfolds and refolds. Our simulations at ≥285 K are in qualitative agreement with the experiment, which have indicated that the GB1 hairpin is ~85% folded at 275 K, ~ 50% folded at 295 K, and ~20% folded at 325 K.[86] However, an anomaly is observed in our 275 K simulation, in which the GB1 hairpin unfolds after ~200 ns and does not refold. This unfolding event may simply be an artifact of our limited sampling that would disappear if the simulations were to run to convergence. Alternatively, it may reflect limitations of the SPC/E_b water model at temperatures distant from those at which it was parametrized; while the temperature-dependent behavior of SPC/E_b has not been characterized, to our knowledge, three-point water models including the parent SPC/E water model are known to poorly reproduce the temperature dependence of properties such as density.[15,87,88]

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

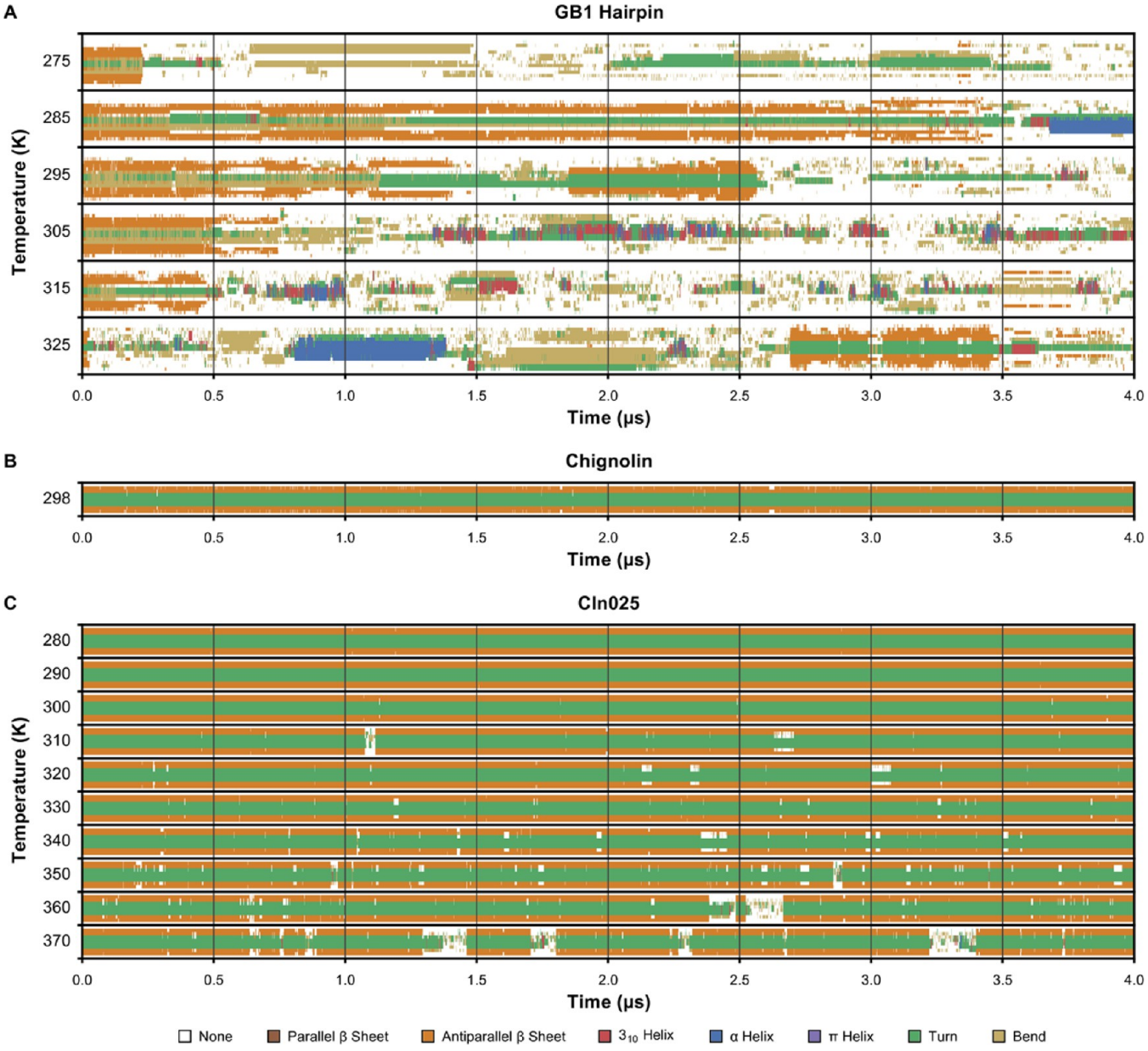


Figure 5. Secondary structures of model β-hairpin peptides ((A) GB1 hairpin, (B) chignolin, and (C) Cln025) at various simulated temperatures over the course of 4-μs simulations.

In contrast, our simulations of chignolin and Cln025 suggest that these β-hairpin systems may be more stable than observed experimentally. As shown in Figure 5B, chignolin maintains its β-hairpin configuration throughout our 4-μs simulation at 298 K, including two hydrogen bonds in an antiparallel sheet configuration, while, experimentally, the peptide is only ~60% folded at this temperature.[54] Chignolin's hyper-stable variant Cln025 has an experimental melting temperature of 343 K.[55] As shown in Figure 5C, in our simulations at temperatures ranging from 280 K to 370 K, we observe unfolding and refolding events at several temperatures, although the overall folded population is larger than that measured experimentally. In particular, Cln025 is >80% folded in our simulation at 370 K, while, experimentally, the peptide is only ~25% folded at this temperature.[55]

As with α-helices, we expect ff15ipq to yield residue-specific propensities in β-sheet stability, although the large difference in sequence between the two tested types of model β-hairpins makes comparing them difficult. The aforementioned lack of a clear β well in Asp may destabilize the GB1 hairpin, which contains two adjacent Asp residues, one of which forms part of the antiparallel β-sheet and the other, the turn. The observed stabilities of chignolin and Cln025 generally preclude the notion that ff15ipq is biased against the β-sheet structure. Further studies including additional hairpin sequences and parallel β-sheet structures will be necessary to quantify and mitigate residue-specific biases for future IPolQ force fields.

4.6. The Trp-Cage Mini-Protein and Globular Proteins BPTI, Villin, GB3, Ubiquitin, Binase, and Lysozyme. In order to assess the stability of proteins with ff15ipq, we simulated the Trp-cage miniprotein and a series of six globular proteins: BPTI, villin, GB3, ubiquitin, binase, and lysozyme. Extensive experimental data are available for all of these model systems, providing excellent opportunities for validation of our

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

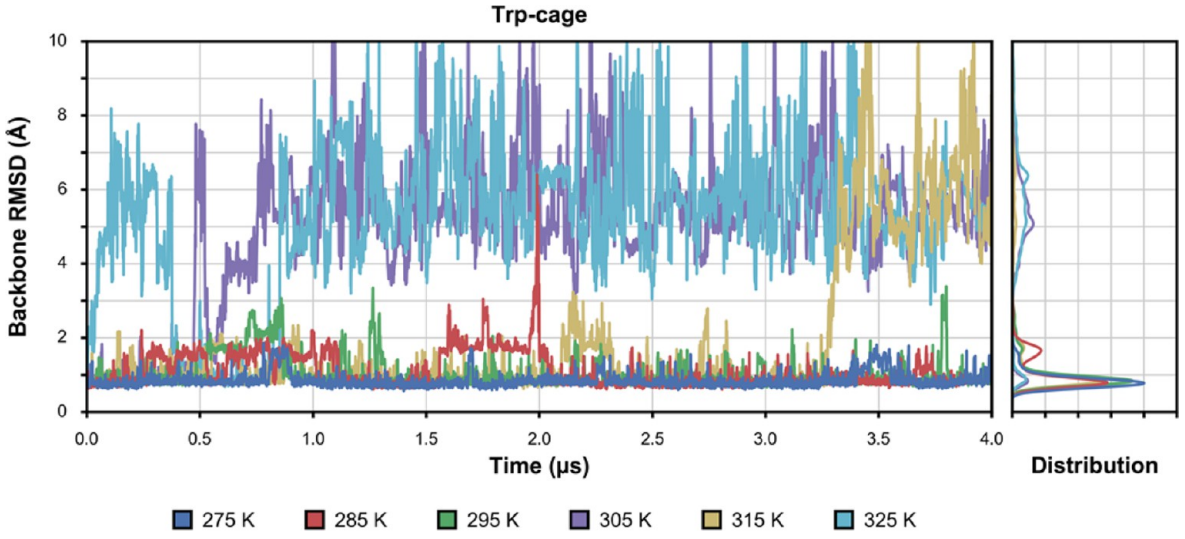


Figure 6. Stability of Trp-cage over the temperature range of 275−325 K over the course of 4-μs simulations, as monitored by the backbone root-mean-square deviation (RMSD), relative to the experimental NMR structure.

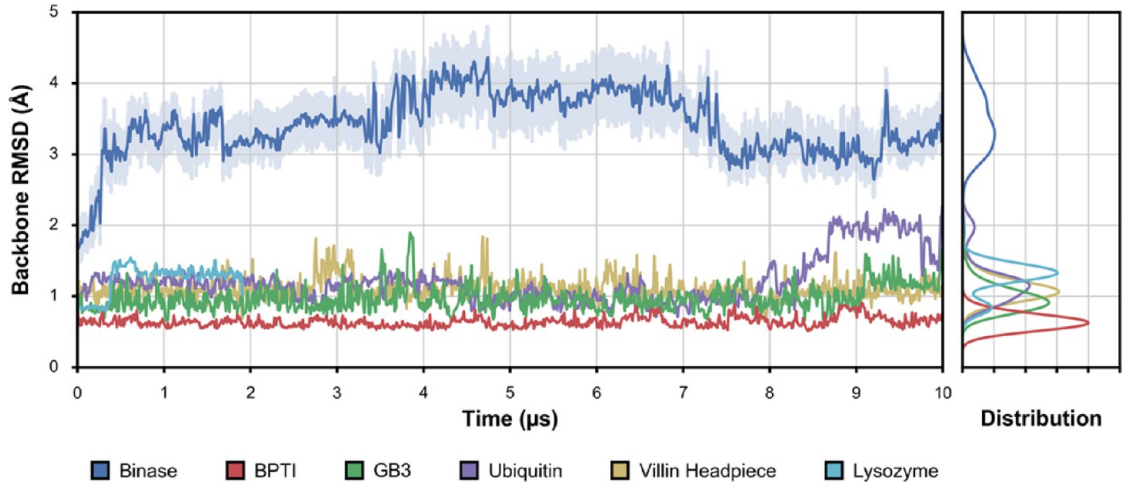


Figure 7. Stability of folded proteins over the course of 10-μs simulations as monitored by the backbone RMSD, relative to the experimental structures. For binase, the mean RMSD relative to the ensemble of 20 NMR structures is shown, along with the range between the minimum and maximum values (light blue shaded region).

simulations. We carried out 24 μs of aggregate equilibrium simulations of Trp-cage at temperatures ranging from 275 K to 325 K and simulations 2−10 μs in duration of the six globular proteins at temperatures between 298 K and 303 K. Details of our simulations are listed in Table 1.

The designed miniprotein Trp-cage is central to a long-running computational success story for AMBER force fields. Folding simulations of this miniprotein using the AMBER ff99SB force field have successfully recovered the folded structure, yielded multiple folding and unfolding events, and provided a melting temperature (T_m) of 318 K, which is in reasonable agreement with the experimental T_m value of ∼315 K.[56,89,90] As shown in Figure 6, Trp-cage remained stable in our simulations between 275 K and 295 K, with an average backbone RMSD from the experimental NMR structure of <1 Å, and unfolded between 305 K and 325 K. While our simulations are not extensive enough to obtain precise estimates of the T_m, these results suggest that the T_m value is somewhere between 295 K and 305 K, which is slightly lower than the experiment. Each unfolding event is marked by an initial shift of the backbone Φ/Ψ of Pro 12 from the α well to the PPII well, followed by the loss of the N-terminal α-helical component of the polypeptide. Notably, in our simulation that was run at 325 K, the protein refolded for ∼500 ns, indicating that the folded state is a stable free-energy minimum. Thus, despite the extensive reoptimization of the parameter set, the important success of the AMBER force fields in modeling the stability of Trp-cage has been maintained.

Note that the Pro-rich sequence of Trp-cage allows the opportunity to validate the unique pre-Pro Ψ and Ψ′ terms of ff15ipq, because it contains Gly, Arg, and Pro residues that precede Pro. Whenever Trp-cage was folded in our simulations, Gly 11 remained stably in its PPII′ well, while Arg 16 was sampled broadly across the β and PPII regions without a clear barrier between them. These results are in good agreement with the experimental NMR ensemble, within which the Φ and Ψ backbone torsions of Arg 16 are distributed in a line across these two regions,[56] and with the pre-Pro distributions observed for these residues in the NDRD dataset.[77] Also consistent with both the NMR ensemble and NDRD dataset

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

are the observed distributions of Pro 17 and Pro 18, which strictly maintained their positions in the PPII well, even as the protein unfolded.

As shown in Figure 7, the overall structures of all six globular proteins—BPTI, villin, GB3, ubiquitin, binase, and lysozyme—remained stable over their entire simulations. The BPTI protein remained closest to its crystal structure, yielding an average backbone RMSD of 0.7 Å, which may be a result of its three disulfide bonds among a total of 58 residues. All α-helical and β-sheet regions of this protein were retained for the entire simulation. The only notable deviation from the crystal structure of BPTI was observed for Ala 16 and Arg 17, which form the end of the loop preceding the first β-sheet. These two residues, which occupy the α and ξ wells, respectively, in the crystal structure, made temporary excursions of several hundred nanoseconds to alternative conformations before returning to the crystal conformation (Figures S5 and S6 in the Supporting Information). The villin headpiece subdomain also remained close to its crystal structure, yielding an average backbone RMSD of 1.1 Å. In order to test our parameters for the noncanonical amino acid norleucine (Nle), which was included alongside the canonical residues during development of ff15ipq, we have simulated the fast-folding double-Nle mutant of villin.[62] The Nle residues, both of which are located in the third α-helix, strictly sampled the α-helical well, suggesting that our parameters for Nle are appropriate. The only significant deviation from the crystal structure of this mutant of villin is a ~0.5 μs excursion made by residues 10−12, which form the end of helix 1 and the loop linking helices 1 and 2, to an alternative conformation different from that observed in the crystal structure (Figures S7 and S8 in the Supporting Information).

Our simulation of GB3 yielded an average backbone RMSD of 1.0 Å from the NMR structure. Three of the residues exhibit significant deviations from the crystal structure: Leu 12, Asp 40, and Thr 55 (Figures S9 and S10 in the Supporting Information). The conformation of Leu 12, which is located in the turn linking the first and second β-strand, falls precisely between the β and PPII wells in the NMR structure, while both wells were almost equally sampled in our simulation. As a result of this increased conformational flexibility at Leu 12, adjacent residues also occasionally sampled conformations outside their NMR structure. The presence of a free-energy barrier between the β and PPII wells is a necessity for maintaining stable conformations within these wells; it is likely that the forces contributing to the stabilization of Leu 12's unusual conformation in the NMR structure are simply not captured by the functional form of ff15ipq. Consistent with this hypothesis, almost-identical deviations were observed for the ff14SB force field that shares this functional form.[4] The conformation of Asp 40, which is located in the loop between the α-helix and third β-strand, occupies the ξ well in the crystal structure while other conformations, predominantly β, were sampled in our simulations. Indeed, based on this result and others presented below, the negatively charged residues of ff15ipq completely lack an ξ well. As with Leu 12, the increased conformational flexibility of Asp 40 led to broader sampling by adjacent residues. It is worth noting that GB3 contains two negatively charged residues, Glu 15 and Asp 22, which remained stable in β-sheets, indicating that the limited sampling of this region observed in our umbrella sampling simulations may be overcome within the context of a folded protein. Finally, a notable deviation from the NMR structure occurs for Thr 55 during the last microsecond of our simulation when the antiparallel β-sheet hydrogen bonds between this residue and Val 42 are broken, although the remainder of the β-sheet remains in place. While this deviation may be a transient event, it could also be a consequence of the conformational deviations of the nearby Asp 40.

Similar to GB3, ubiquitin in our simulations exhibited a low overall average backbone RMSD of 1.2 Å from the crystal structure, but significant deviations in certain regions. In particular, transient deviations from the crystal structure were observed for residues 8−11, which form the turn connecting the first and second β-strands (see Figures S12 and S13 in the Supporting Information). While these residues sampled conformations that were different from the crystal structure, their turn conformation was retained throughout 80% of the simulation, and experimental NMR relaxation data suggests that the region is truly flexible.[91] A more significant deviation was observed for Asp 52 and Gly 53, which are located in a loop region; in the crystal structure, these residues both occupy the α well, whereas, in our simulation, Asp 52 and Gly 53 shift to the PPII and γ wells, respectively. Unlike the deviations observed in GB3, this shift does not lead to broader sampling by adjacent residues or appear to otherwise destabilize the protein, suggesting that the observed alternative conformation may simply be erroneously modeled by ff15ipq to be lower in energy than the conformation found in the crystal structure. Finally, after ~8.5 μs of simulation, Glu 34, which is the last helical residue in the central α-helix, shifts to the PPII and β regions, leading to shifts in residues 33−36. Combined with the observations made for Asp 52, this shift suggests that, for negatively charged residues, ff15ipq may overstabilize PPII conformations, relative to α.

Among the simulated globular proteins, the greatest deviations from the initial structure were observed for binase with an average backbone RMSD of 3.4 Å from the NMR ensemble of 20 models.[57] The same average RMSD was also obtained, with respect to the crystal structure of wild-type binase, which differs in the amino acid sequence at six positions (PDB code: 1GOU).[92] These larger differences are primarily caused by variability in loop regions, as noted for the experimental structures,[57,92] and the core structure of binase remained relatively close to the experimental structures with an average backbone RMSD of 1.9 Å. The first loop, which consists of residues 34−39, adopted multiple conformations in our simulation, which is consistent with the NMR ensemble (Figures S15 and S16 in the Supporting Information). Moreover, this loop adopts the conformation observed in the crystal structure for 75% of the simulation. The second loop consists of residues 56−62, which also sampled broadly in our simulations, consistent with diverse conformations in the NMR ensemble and poorly defined electron density in the crystal structure.[92] Notably, in both our simulation and the NMR ensemble, flexibility in this region extends to Gly 67. This difference may relate to the difference in sequence between the NMR and crystal structure proteins, in which Ser 66 is replaced by Gly and Gly 67 by Ser. The third loop, comprised of residues 76−83, is also broadly sampled in our simulation, and is the source of greatest difference, relative to the experimental structures. The final loop is comprised of residues 99−104 and, in our simulation, we observe several residues sampling two different conformations, consistent with the observation of two states in crystal structures determined under different conditions.[92]

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

The largest protein system, lysosome, was simulated for 2 μs, over which it exhibited an average backbone RMSD of 1.2 Å, with respect to the crystal structure. As shown in Figures S17 and S18 in the Supporting Information, the largest deviations were found in the loop comprised of residues 100−104. Residues 101−104 adopted an alternative conformation, with the flanking residue Val 99 no longer part of the preceding α helix and Gly 104 part of the following helix. Similar to our observations for ubiquitin, this difference appears to be related to the shift of a negatively charged residue, Asp 101, from the α well to the PPII well.

Four of the six globular proteins—BPTI, GB3, ubiquitin, and lysozyme—have also been used for validation of previously developed force fields, including ff14ipq, ff14SB, and CHARMM36.[4,11,17] Similar to these force fields, ff15ipq yielded low average backbone RMSD values for these proteins, relative to their initial structures (i.e., ≤1.2 Å). A key point to be considered while comparing our results to those of previous force fields is that advancements in GPU computing over the last several years[2] have enabled us to validate ff15ipq using simulations up to 10 μs, which is up to 10 times longer than those used for the other force fields. In particular, ff14SB was validated using sets of four 1-μs simulations,[4] while CHARMM36 was validated using 200 ns simulations.[11] Many of the key deviations that we observe do not occur for several microseconds; for example, we observe changes in the C-terminus of GB3 after 9 μs, and in the loops of ubiquitin after 8.5 μs. These deviations are informative and will guide development of successors to ff15ipq, illustrating the utility of long-time scale simulations for force field development.

A particularly appealing feature of the SPC/E_b water model with which we have developed ff15ipq is its ability to more accurately reproduce the rotational diffusion of solvated proteins relative to water models, such as TIP3P and TIP4P-Ew.[23] In order to measure how accurately the combination of ff15ipq and SPC/E_b are able to reproduce rotational diffusion, we branched off sets of ten 200-ns simulations in the microcanonical ensemble (NVE) from our 10-μs simulations in the canonical ensemble (NVT) for GB3, ubiquitin, and binase, thereby avoiding perturbation of the dynamics by the use of a thermostat. As shown in Table 3, ubiquitin, GB3, and binase diffused ∼14%, ∼15%, and ∼22% more slowly than that measured experimentally by NMR, respectively. Note that the experimental values were corrected for differences in temperature, isotopic labeling, and solvent D_2O content, potentially introducing error into the comparison of simulated and experimental values.[29] Interestingly, while the errors that we obtained were consistent with our test simulations with the AMBER ff99SB-ILDN force field and SPC/E_b (see Figure S1), we found CHARMM22* and SPC/E_b to yield lower errors of 7%, 6%, and 16%, illustrating the coupling of solute and solvent parameters on the motions of proteins through solution. Since the SPC/E_b water model had been empirically optimized for proteins with the AMBER ff99SB force field, this is no fault of ff15ipq (or CHARMM22*, for that matter), but suggests that improved performance might be obtained by optimizing the protein and solvent models in tandem. Finally, it is worth noting that our use of the Langevin thermostat in the NVT simulations results in ∼42%−52% longer rotational diffusion times, relative to those of the NVE simulations, demonstrating, for the first time (to our knowledge), that the use of a thermostat can significantly perturb dynamical properties for proteins and not just for small molecules and polymer chains.[93]

Table 3. Rotational Diffusion of Globular Proteins Simulated with ff15ipq[a]

system	experimental τ_c (ns)[b]	Simulated τ_c (ns)	
		one 10-μs NVT simulation[c]	ten 200-ns NVE simulations[d]
GB3	3.03	4.94 ± 0.02	3.47 ± 0.05
ubiquitin	4.07	6.67 ± 0.04	4.62 ± 0.08
binase	5.95	11.06 ± 0.07	7.26 ± 0.18

[a]Uncertainties represent one standard error of the mean. [b]Experimental rotational diffusion measured using NMR relaxation[59,91,109] and corrected for differences in temperature and D_2O content between simulation and experiment.[29] [c]Uncertainties represent one standard error of the mean calculated from 50 consecutive 200-ns blocks from a single 10-μs simulation. [d]Uncertainties represent one standard error of the mean calculated from ten independent 200-ns simulations.

A major advantage of performing simulations with accurate rotational diffusion is that one can directly calculate NMR relaxation parameters ^{15}N R_1 and R_2, and the $^{15}N-^1H$ heteronuclear NOE, which report on the dynamics of individual residues, and compare these values with experiment. Therefore, we calculated relaxation parameters for GB3 and ubiquitin, for which experimental data are available at five and four magnetic field strengths, respectively (see Figures S11 and S14, respectively, in the Supporting Information). Overall, our calculated R_2 values are in excellent agreement with the experiment, with average mean absolute percent error (MAPE) values of 8% for GB3 and 9% for ubiquitin. Our R_1 values are also in good agreement, with average MAPE values of 10% and 12%, with a consistent offset observed across all residues. However, our heteronuclear NOE values are somewhat poorer agreement, with average MAPE values of 22% and 30% for the two systems.

Residues for which our calculated R_1 and R_2 differ by >20% from the experimental values may be due to limitations of our force field. Notably, we find that several such residues are those for which we also observed deviation in sampled backbone conformations, relative to the experimental structures. For GB3, Leu 12 and Asp 40 yielded above-average errors in R_1 (25%), while Gly 41 had a larger error in both R_1 and R_2 (47% and 33%). Similarly for ubiquitin, Lys 11 and Asp 52 yielded above average errors in R_1 (20%). Also in ubiquitin, Asn 25 yielded an error of 22% in R_2; however, this residue is found experimentally to undergo chemical exchange,[91] fluctuating at time scales beyond those captured by our simulations. Among residues with errors below 20%, one particular trend is apparent: Ile 7, Thr 16, Thr 49, and Thr 51 of GB3 and Thr 12 and Ile 36 of ubiquitin all yielded errors in R_2 of ≥15%, despite having tightly restricted Φ/Ψ sampling consistent with their experimental structures. This trend suggests that ff15ipq may have some discrepancy with the three branched residues that is not apparent when examining only backbone Φ/Ψ preferences, and this will be the subject of further study.

4.7. Disordered Peptides: p53 Peptide, S-Peptide. In order to test the suitability of ff15ipq for simulating disordered proteins, we focused on two model peptides: the N-terminal, 13-residue peptide fragment of the tumor suppressor p53, and the 22-residue S-peptide fragment of RNase A. Both of these disordered peptides (p53 peptide and S-peptide) only adopt α-helical conformations when bound to their structured partner proteins (MDM2 and S-protein, respectively).[94,95] For each of these peptides, we carried out two 10-μs simulations: one of the

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

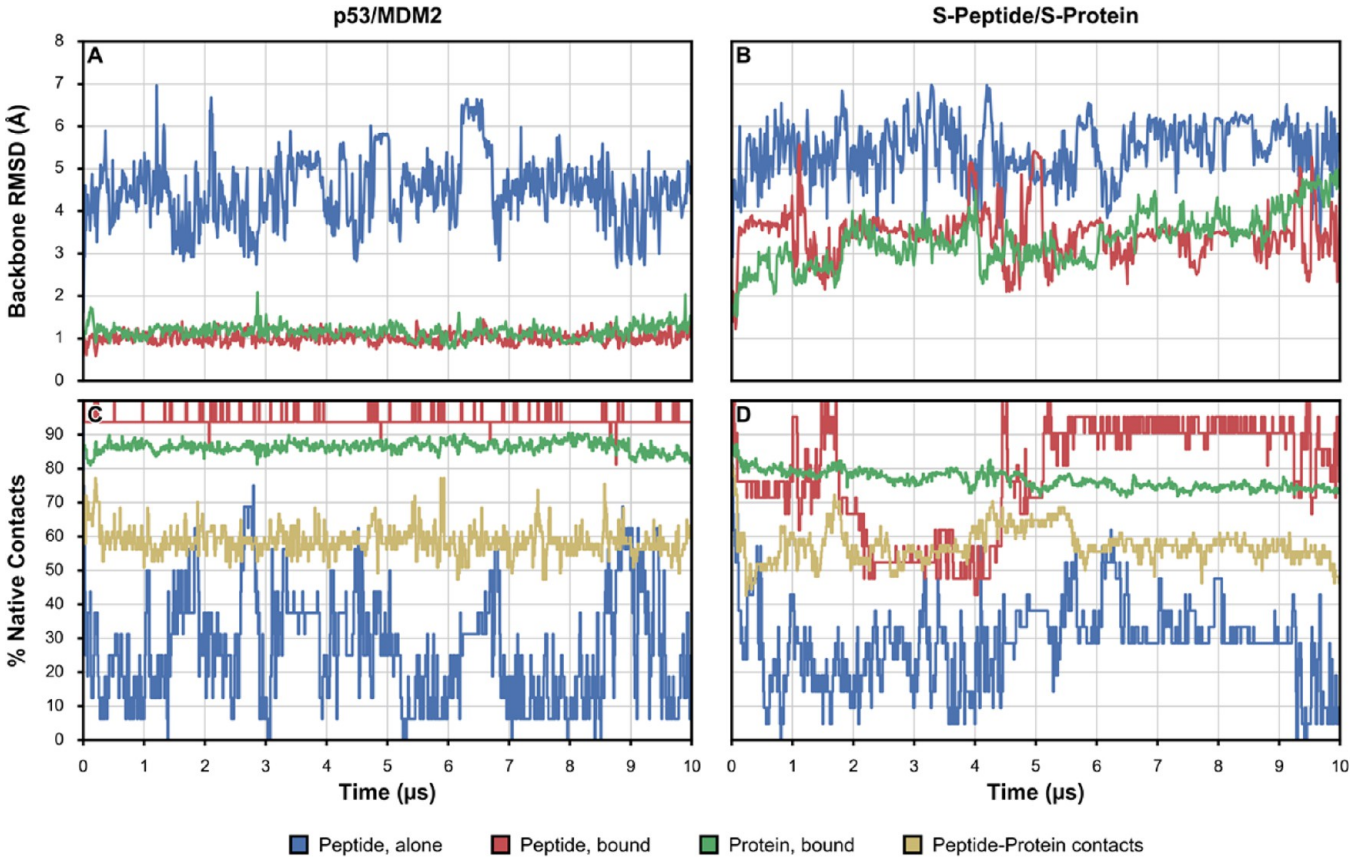


Figure 8. Stability of p53 and S-peptide alone and in complex with binding partners MDM2 and S-protein over the course of 10-μs simulations as measured by backbone RMSD relative to (A, B) the crystal structures and (C, D) the percent of native contacts formed.

isolated peptide and the other of the native peptide−protein complex.

As shown in Figure 8, both peptides in their isolated states adopted conformations distant from their partner-bound conformations, sampling a diverse set of conformations with average backbone RMSDs of 5 Å from their corresponding bound conformations in the crystal structures of the peptide−protein complexes and maintaining only ∼25% of their native intrapeptide contacts. Furthermore, the p53 peptide only transiently adopted α-helical conformations that resembled its partner-bound conformations (backbone RMSD value of <3 Å) with these conformations unfolding within ∼200 ns (see Figure S19 in the Supporting Information). In contrast, the S-peptide did not even transiently sample α-helical conformations resembling its bound state; instead, the peptide formed β-hairpins that persisted for periods as long as 4 μs (see Figure S22 in the Supporting Information).

In our simulation of the p53/MDM2 complex, the peptide remained stably bound to its partner protein for the entire 10 μs of simulation, with an average backbone RMSD of 1 Å from its bound conformation in the crystal structure (Figure 8). Curiously, while most of the intramolecular native contacts of the p53 peptide and MDM2 were retained (∼95% and ∼85%, respectively), only ∼60% of the intermolecular p53/MDM2 native contacts persisted. Examination of the structure showed that many of these contacts lay just below the threshold distance of 5.5 Å between their heavy atoms in the crystal structure. Thus, these contacts were no longer "formed" when slightly different conformations were adopted in our simulations. Throughout our simulation, the p53 peptide retained the α-helical structure of residues 19−25 (recall Figure S19). In contrast, our simulation of the S-peptide/S-protein complex sampled conformations more distant from its crystal structure, with average backbone RMSD values of 3 Å for both S-peptide and S-protein. Although the S-peptide partially lost its α-helical structure near the N-terminus from ∼1.5 μs to ∼5 μs in our simulation (see Figure 8 and Figure S22), the structure reformed and persisted for the remainder of the simulation during which both the S-peptide and S-protein retained most of their intramolecular native contacts (∼90% and ∼80%, respectively). Similar to the p53/MDM2 complex, the S-peptide/S-protein complex retained ∼60% of its peptide−protein contacts throughout our simulation.

Our simulations of these flexible, disordered peptides provide an additional opportunity to validate the backbone conformational preferences of ff15ipq. From our four simulations, we have calculated the backbone Φ/Ψ sampling of the p53 peptide (Figure 9) and S-peptide (Figure S23 in the Supporting Information). For comparison, we have provided the distributions for the peptide sequences obtained from the NDRD dataset, which accounts for the influence of adjacent residues on each distribution.[77] In our simulation of the p53/MDM2 complex, residues 18−27 of p53 occupied exclusively the wells observed in the crystal structure, aligning with the observed low RMSD and high percentage of native contacts. Our simulation of the isolated p53 peptide exhibits similarities with the NDRD distribution for most residues, but several inconsistencies are informative: in particular, the neutral residues exhibit reasonable overall agreement and the bulky aromatic residues Phe 19 and Trp 23 closely resemble the

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

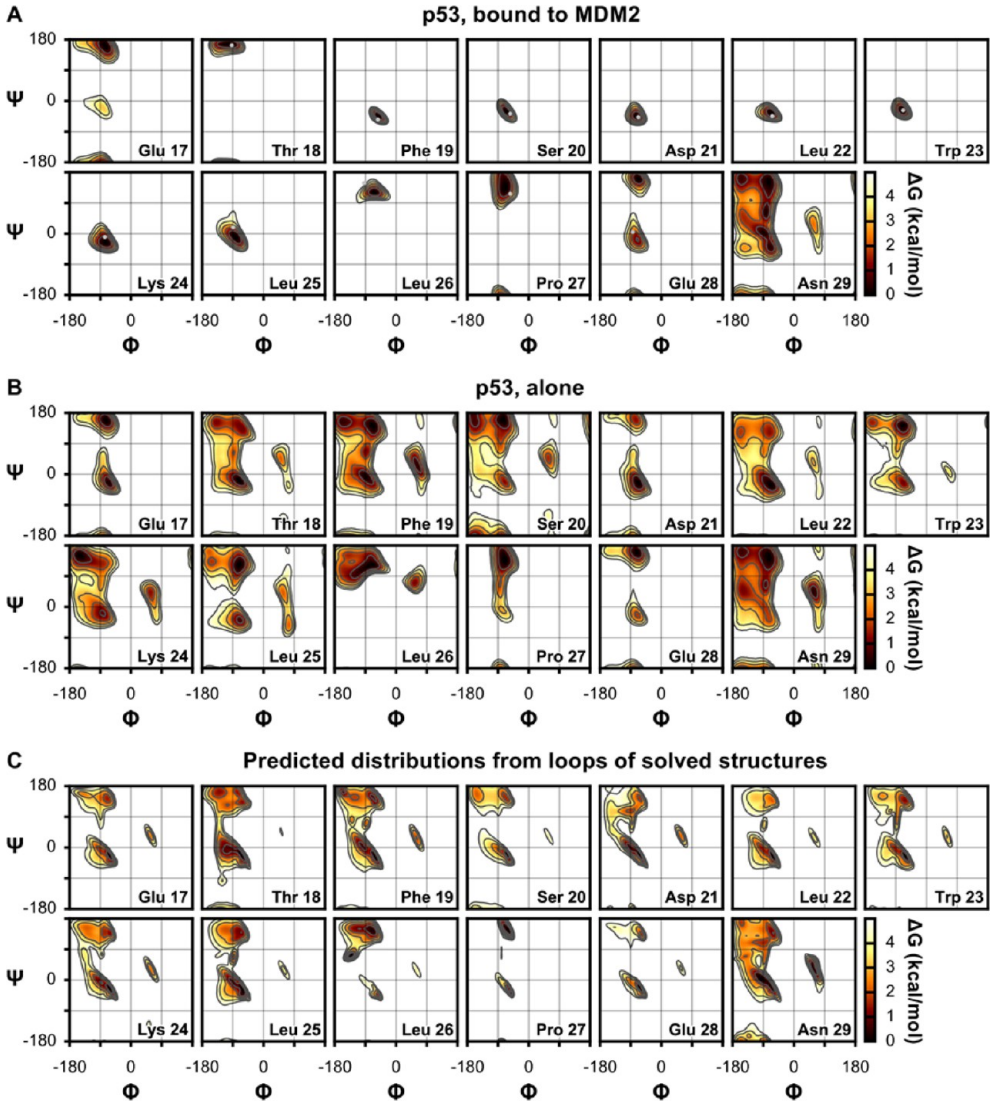


Figure 9. Backbone conformational sampling of the disordered p53 peptide observed in 10-μs simulations (A) in complex with the MDM2 protein and (B) alone. For comparison, panel (C) shows distributions for the p53 sequence obtained from the Neighbor-Dependent Ramachandran Distribution (NDRD) dataset,[77] derived from conformations observed in the loops of crystal structures.[77]

NDRD distributions, while Leu 22 and Leu 25 sampled of the β-region more extensively than suggested by the NDRD. The negatively charged residues Glu 17 and Glu 28 sampled consistently with the NDRD, while Asp 21 missed sampling in the ξ region. The sole positively charged residue, Lys 24, sampled the β-well more extensively than suggested by the NDRD.

In our simulation of the S-peptide/S-protein complex, the crystal conformations were retained for most of the simulation (see Figure S23). As described above, several residues near the N-terminus left the α-helical well, and the adjacent Thr 3, which is not helical in the crystal structure, eventually joined the helix as it reforms. The clearest difference from the crystal structure is found for Asp 14; this residue occupies the ξ well in the crystal structure, which is not present for Asp in ff15ipq, causing it to adopt a PPII conformation. Within the S-protein, residue Gln 60 is notable for its uncommon "plateau" conformation ($\Phi \approx -100°$, $\Psi \approx -130°$),[64,96] which was retained throughout our simulation (Figure S25 in the Supporting Information). In our simulation of the isolated, unbound S-peptide, the formation of long-lived β-hairpin structures prevented us from obtaining converged conformational preferences for comparison with the NDRD distributions, despite the long duration of the simulations (10 μs).

Taken together, the above results indicate that ff15ipq can reliably predict disorder as well as order for peptides that fold upon binding their partner proteins. These encouraging results are worth pointing out since ff15ipq was not specifically parametrized for disordered peptides/proteins, as is the case for contemporary force fields such as ff03w and its subsequent variants.[9,97] As shown in Figure 1, both ff03 (whose atomic charges and radii are shared by ff03w) and ff15ipq are able to reliably model propensities of salt-bridge formation, which can be critical for such systems that are rich in polar and/or charged residues. Thus, ff15ipq is a reasonable alternative to ff03w for the simulation of disordered peptides/proteins.

5. DISCUSSION

Since the establishment of the "one atom, one site" model of all-atom fixed-charge force fields over 20 years ago, several major lineages of protein force fields and countless branches have been developed through cycles of validation and

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

refinement. In this work, we present the ff15ipq force field, the latest in the AMBER IPolQ lineage, and validate its accuracy by >200 μs of aggregate MD simulation. The distinguishing features of ff15ipq are (i) a charge set that accounts directly for water induced polarization, (ii) the incorporation of two related charge sets for creating new force fields based purely on *ab initio* calculations, (iii) the scope of the parameter optimization, including backbone angles alongside torsions, and (iv) the degree of automation and transferability of the methods to other regions of chemical space. Our simulations suggest that ff15ipq yields reasonable salt-bridge propensities, maintains secondary structures and globular protein folds on the μs time scale, predicts order as well as disorder in protein structures, and yields strong agreement with NMR J-couplings and relaxation rates. However, even with this extensive amount of validation, several unconverged results remain and will be explored further using enhanced sampling techniques such as replica exchange[98,99] or recent variants[100,101] of the weighted ensemble path sampling strategy.[102] A timeline of the development is available in the Supporting Information. Here, we will discuss the origins of ff15ipq and its current trajectory.

A major motivation for creating ff15ipq was to address concerns about the overstabilization of salt-bridge interactions by ff14ipq, which is a limitation that is shared by other contemporary force fields.[19] We addressed these concerns by abandoning the mixed Lennard-Jones radii of ff14ipq and instead increasing the radii of polar hydrogen atoms bonded to nitrogens in both the protein backbone and side-chains to yield more accurate salt-bridge propensities. The atomic charges of the force field were then rederived by applying the automated machinery, which had created ff14ipq, exchanging the TIP4P-Ew water model for SPC/E_b. Finally, all torsion and selected backbone angle parameters were refit to a QM dataset over four times as large as that used for ff14ipq. In doing so, we found that angle optimization was essential for recovering reliable results with our more-extensive dataset (see the "Timeline of Development" section in the Supporting Information).

While the angle optimization feature and other particulars of the torsion fitting are subjects of ongoing development in our force field engine mdgx, our results with ff15ipq indicate that the workflow illustrated in Figure 10 is a viable approach to creating new force fields. Each step entails additional layers of details in order to address practical considerations such as infinite electrostatics or the forms of molecular mechanics basis functions and the shape of the target energy surface. We have addressed each of these issues in the Theory and Methods sections as well as in prior publications,[16,17] but the synthesis of all these details reflects the physical arguments behind the IPolQ charge derivation.

This approach should be viewed in context with the contemporary Force Balance approach, which also performs sweeping optimization of hundreds of parameters simultaneously.[15] Unlike our MM-minimized conformations, Force Balance typically considers conformations QM-minimized at the same level of theory at which the target QM energies are calculated, although this is not a strict requirement. Going beyond the capabilities of mdgx, Force Balance includes numerous nonlinear optimization methods and offers the capability to incorporate results from sources beyond QM single-point energies, including *in vitro* experiments, directly into the parameter optimization. In the future, such diverse targets might be paired with the IPolQ method, for example, by using the vacuum charge set (Q_{vac}) for comparison with vacuum-phase QM data, but using the polarized charge set (Q_{solv}) in simulations for comparison to experimental results in solution.

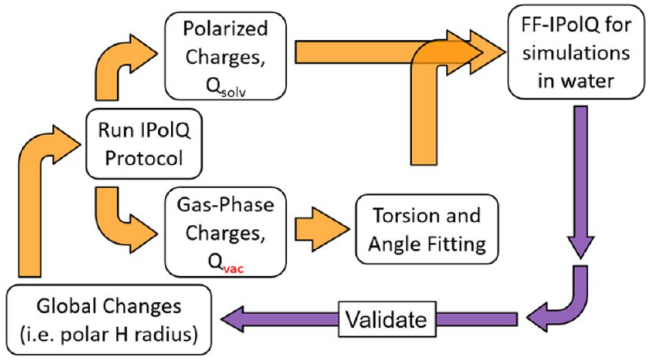


Figure 10. IPolQ force field development workflow. Starting from an existing model, selected global changes are optionally first applied to obtain an initial model for optimization. The IPolQ charge deviation protocol is then used to fit a pair of atomic charge sets for the vacuum (Q_{vac}) and solution (Q_{solv}) phases. The vacuum-phase charges are used to fit parameters for bonded terms to vacuum-phase QM targets, and these parameters are subsequently paired with the solution-phase charges to yield a complete force field for solution-phase simulations. The force field is then validated through extensive MD simulation, informing future development.

It is rather remarkable that a viable protein force field can be produced within months, almost entirely from QM data. Also noteworthy is the fact that features such as angle optimization and generational refinement, which had incremental but definite effects on the accuracy of data fitting, could be so influential in the final result. One way of considering the remaining error in our MM model is to partition it between two sources: bonded and nonbonded interactions. The improvements in ff15ipq that were obtained relative to ff14ipq, whose nonbonded parameters are of similar accuracy, resulted from optimization of angles and the branching of bonded parameters. While the inclusion of anharmonicity in bond and angle stretching or a spline-based treatment of torsion cross terms (CMAP) may reduce errors further,[6,103] greater improvements might be accomplished in the nonbonded interactions.

For this reason, the next planned advance of the AMBER IPolQ force field lineage is to improve the accuracy of electrostatic interactions by making liberal use of virtual charge sites. The "one atom, one site" paradigm used for ff15ipq, which was established several decades ago, appears sufficiently accurate for most purposes, economical by construction, and thoroughly optimized in existing MD engines. Models with significant numbers of virtual charge sites are presently in the process of becoming established. These models offer improved accuracy for various chemistries with clear physical motivations, accompanied by a more modest increase in computational cost than that afforded by polarizable functional forms.[104,105] Future AMBER IPolQ development at the one-site-per-atom level will continue to explore the applicability of our methodology to the chemical space of other biologically important molecules, including nucleic acids, carbohydrates, and small molecules. Further ahead along the path lie multisite IPolQ models, which will push the mimicry of QM potential energy surfaces—within the confines of a nonpolarizable model—to new levels.

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00567.

Timeline of the development of ff15ipq, including analysis and testing of development versions of the force field (PDF)

Methods for simulations run using polarizable force fields and on the Anton supercomputer, the results of tests used to select the water model and hydrogen radius for ff15ipq, and detailed analysis of validation simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +1 412 624-6026. Fax: +1 412 624-8301. E-mail: ltchong@pitt.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by a University of Pittsburgh Andrew Mellow Fellowship (to K.T.D.), NIH Grant No. 1RO1GM115805-01 (to L.T.C.), NIH Grant No. GM45811 (to D.A.C.), and NIH Grant No. RO1GM080642 (to A.M.G.). Computational resources were provided through NSF XSEDE Award No. MCB-100109 for the use of TACC's Stampede; D.E. Shaw Research, NRBSC, PSC, and BTRC for Multiscale Modeling of Biological Systems (MMBioS) through NIH grant P41GM103712-S1 for the use of Anton; NSF MRI Award No. CNS-1229064 for a shared cluster at the University of Pittsburgh's Center for Simulation and Modeling; and private clusters in the Case Laboratory.

■ REFERENCES

(1) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. GPU-Accelerated Molecular Modeling Coming of Age. *J. Mol. Graphics Modell.* **2010**, *29*, 116−125.

(2) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184* (2), 374−380.

(3) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.

(4) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(5) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78* (8), 1950−1958.

(6) Li, D.-W.; Brüschweiler, R. Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *J. Chem. Theory Comput.* **2011**, *7* (6), 1773−1782.

(7) Li, D.-W.; Brüschweiler, R. NMR-Based Protein Potentials. *Angew. Chem., Int. Ed.* **2010**, *49* (38), 6778−6780.

(8) Zhou, C.-Y.; Jiang, F.; Wu, Y.-D. Residue-Specific Force Field Based on the Protein Coil Library. RSFF2: Modification of AMBER ff99SB. *J. Phys. Chem. B* **2015**, *119*, 1035−1047.

(9) Best, R. B.; Mittal, J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010**, *114* (46), 14916−14923.

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

(11) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Φ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

(12) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100* (9), L47−L49.

(13) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.

(14) Jiang, F.; Zhou, C. Y.; Wu, Y. D. Residue-Specific Force Field Based on the Protein Coil Library. RSFF1: Modification of OPLS-AA/L. *J. Phys. Chem. B* **2014**, *118* (25), 6983−6998.

(15) 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.

(16) 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. *J. Phys. Chem. B* **2013**, *117*, 2328−2338.

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

(18) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Götz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P. A.; Kaus, J. W.; Kovalenko, A.; Lee, T.; Le Grand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B. D.; Merz, K. M.; Monard, G.; Needham, H.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A. E.; Salomon-Ferrer, R.; Simmerling, C.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*; University of California: San Francisco, CA, 2015.

(19) Debiec, K. T.; Gronenborn, A. M.; Chong, L. T. Evaluating the Strength of Salt Bridges: A Comparison of Current Biomolecular Force Fields. *J. Phys. Chem. B* **2014**, *118*, 6561−6569.

(20) Xue, Y.; Yuwen, T.; Zhu, F.; Skrynnikov, N. R. Role of Electrostatic Interactions in Binding of Peptides and Intrinsically Disordered Proteins to Their Folded Targets. 1. NMR and MD Characterization of the Complex between the c-Crk N-SH3 Domain and the Peptide Sos. *Biochemistry* **2014**, *53* (41), 6473−6495.

(21) Yuwen, T.; Xue, Y.; Skrynnikov, N. R. Role of Electrostatic Interactions in Binding of Peptides and Intrinsically Disordered Proteins to Their Folded Targets. 2. The Model of Encounter Complex Involving the Double Mutant of the c-Crk N-SH3 Domain and Peptide Sos. *Biochemistry* **2016**, *55*, 1784−1800.

(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) Takemura, K.; Kitao, A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. *J. Phys. Chem. B* **2012**, *116*, 6279−6287.

(24) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(25) Karamertzanis, P. G.; Raiteri, P.; Galindo, A. The Use of Anisotropic Potentials in Modeling Water and Free Energies of Hydration. *J. Chem. Theory Comput.* **2010**, *6*, 1590−1607.

(26) 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.* **1993**, *97*, 10269−10280.

(27) Hu, H.; Lu, Z.; Yang, W. Fitting Molecular Electrostatic Potentials from Quantum Mechanical Calculations. *J. Chem. Theory Comput.* **2007**, *3* (3), 1004−1013.
(28) Jarymowycz, V. A.; Stone, M. J. Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chem. Rev.* **2006**, *106*, 1624−1671.
(29) Wong, V.; Case, D. A. Evaluating Rotational Diffusion from Protein MD Simulations. *J. Phys. Chem. B* **2008**, *112*, 6013−6024.
(30) 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.
(31) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(32) Vanommeslaeghe, K.; Yang, M.; MacKerell, A. D., Jr. Robustness in the Fitting of Molecular Mechanics Parameters. *J. Comput. Chem.* **2015**, *36*, 1083−1101.
(33) Hopkins, C. W.; Roitberg, A. E. Fitting of Dihedral Terms in Classical Force Fields as an Analytic Linear Least-Squares Problem. *J. Chem. Inf. Model.* **2014**, *54* (7), 1978−1986.
(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197.
(35) Ho, B. K.; Brasseur, R. The Ramachandran Plots of Glycine and Pre-Proline. *BMC Struct. Biol.* **2005**, *5*, 14.
(36) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr. Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9* (12), 5430−5449.
(37) Shi, Y.; Xia, Z.; Zhang, J.; Best, R. B.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046−4063.
(38) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J. C.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283−312.
(39) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr.; Schulten, K.; Roux, B. High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *J. Phys. Chem. Lett.* **2011**, *2*, 87−92.
(40) Springs, B.; Haake, P. Equilibrium Constants for Association of Guanidinium and Ammonium Ions with Oxyanions. *Bioorg. Chem.* **1977**, *6* (2), 181−190.
(41) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA Polarizable Force Field Parameterization for Small Molecules. *Theor. Chem. Acc.* **2012**, *131*, 1138.
(42) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91* (1), 461−466.
(43) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum through Argon Revisited. *J. Chem. Phys.* **2001**, *114* (21), 9244−9253.
(44) Woon, D. E.; Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98* (2), 1358−1371.
(45) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96* (9), 6796−6806.
(46) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007−1023.
(47) Neese, F.; Wennmohs, F.; Becker, U.; Bykov, D.; Ganyushin, D.; Hansen, A.; Izsák, R.; Liakos, D. G.; Kollmar, C.; Kossmann, S.; Pantazis, D. A.; Petrenko, T.; Reimann, C.; Riplinger, C.; Roemelt, M.; Sandhöfer, B.; Schapiro, I.; Sivalingam, K.; Wezisla, B.; Kállay, M.; Grimme, S.; Valeev, E.; Chan, G. *Orca 3.0.3*, 2015; available via the Internet at: https://orcaforum.cec.mpg.de/.
(48) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 383−394.
(49) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995**, *16* (11), 1339−1350.
(50) Grossfield, A. *WHAM: The Weighted Histogram Analysis Method. Version 2.0.9*; available via Internet at:http://membrane.urmc.rochester.edu/content/wham.
(51) 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.
(52) Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of Helicity within the Model Peptide Acetyl(AAQAA)3amide. *J. Am. Chem. Soc.* **1994**, *116* (5), 8288−8293.
(53) Blanco, F. J.; Rivas, G.; Serrano, L. A Short Linear Peptide That Folds into a Native Stable β-Hairpin in Aqueous Solution. *Nat. Struct. Biol.* **1994**, *1* (9), 584−590.
(54) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. 10 Residue Folded Peptide Designed by Segment Statistics. *Structure* **2004**, *12* (8), 1507−1518.
(55) 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. *J. Am. Chem. Soc.* **2008**, *130* (46), 15327−15331.
(56) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9* (6), 425−430.
(57) Reibarkh, M. Y.; Nolde, D. E.; Vasilieva, L. I.; Bocharov, A. A.; Shulga, A. A.; Kirpichnikov, M. P.; Arseniev, A. S. Three-Dimensional Structure of Binase in Solution. *FEBS Lett.* **1998**, *431*, 250−254.
(58) Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. Structure of Bovine Pancreatic Trypsin Inhibitor. Results of Joint Neutron and X-ray Refinement of Crystal Form II. *J. Mol. Biol.* **1984**, *180*, 301−329.
(59) 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. *J. Am. Chem. Soc.* **2003**, *125* (30), 9179−9191.
(60) Walsh, M. A.; Schneider, T. R.; Sieker, L. C.; Dauter, Z.; Lamzin, S.; Wilson, K. S. Refinement of Triclinic Hen Egg-White Lysozyme at Atomic Resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 522−546.
(61) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194* (3), 531−544.
(62) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J. Sub-Microsecond Protein Folding. *J. Mol. Biol.* **2006**, *359* (3), 546−553.
(63) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science* **1996**, *274*, 948−953.
(64) Kim, E. E.; Varadarajan, R.; Wyckoff, H. W.; Richards, F. M. Refinement of the Crystal Structure of Ribonuclease S. Comparison with and between the Various Ribonuclease A Structures. *Biochemistry* **1992**, *31* (49), 12304−12314.
(65) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with Amber—Part I: Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.
(66) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: New York, 1989.
(67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577−8593.

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

(68) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *N*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(69) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952−962.
(70) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864−1874.
(71) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084−3095.
(72) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(73) Prompers, J. J.; Brüschweiler, R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2002**, *124* (16), 4522−4534.
(74) Veenstra, D. L.; Ferguson, D. M.; Kollman, P. A. How Transferable Are Hydrogen Parameters in Molecular Mechanics Calculations? *J. Comput. Chem.* **1992**, *13* (8), 971−978.
(75) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586−3616.
(76) Tanford, C. The Association of Acetate with Ammonium and Guanidinium Ions. *J. Am. Chem. Soc.* **1954**, *76*, 945−946.
(77) Ting, D.; Wang, G.; Shapovalov, M. V.; Mitra, R.; Jordan, M. I.; Dunbrack, R. L., Jr. Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. *PLoS Comput. Biol.* **2010**, *6* (4), e1000763.
(78) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank. *Eur. J. Biochem.* **1977**, *80* (2), 319−324.
(79) Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J.* **2008**, *95* (1), L07−L09.
(80) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. *J. Am. Chem. Soc.* **2007**, *129* (5), 1179−1189.
(81) Hennig, M.; Bermel, W.; Schwalbe, H.; Griesinger, C. Determination of ψ Torsion Angle Restraints from 3J(Cα,Cα) and 3J(Cα,HN) Coupling Constants in Proteins. *J. Am. Chem. Soc.* **2000**, *122* (26), 6268−6277.
(82) Wirmer, J.; Schwalbe, H. Angular Dependence of 1J(Ni,Cαi) and 2J(Ni,Cα(i-1)) Coupling Constants Measured in J-Modulated HSQCs. *J. Biomol. NMR* **2002**, *23* (1), 47−55.
(83) Ding, K.; Gronenborn, A. M. Protein Backbone 1HN-13Cα and 15N-13Cα Residual Dipolar and J Couplings: New Constraints for NMR Structure Determination. *J. Am. Chem. Soc.* **2004**, *126*, 6232−6233.
(84) Case, D. A.; Scheurer, C.; Brüschweiler, R. Static and Dynamic Effects on Vicinal Scalar J Couplings in Proteins and Peptides: A MD/DFT Analysis. *J. Am. Chem. Soc.* **2000**, *122*, 10390−10397.
(85) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. Interpreting Dynamically-Averaged Scalar Couplings in Proteins. *J. Biomol. NMR* **2005**, *32* (4), 273−280.
(86) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding Dynamics and Mechanism of β-Hairpin Formation. *Nature* **1997**, *390*, 196−199.
(87) Sanz, E.; Vega, C.; Abascal, J. L. F.; MacDowell, L. G. Phase Diagram of Water from Computer Simulation. *Phys. Rev. Lett.* **2004**, *92* (25), 1−4.
(88) Vega, C.; Sanz, E.; Abascal, J. L. F. The Melting Temperature of the Most Common Models of Water. *J. Chem. Phys.* **2005**, *122* (11), 114507.
(89) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A. E.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 712−725.
(90) Paschek, D.; Day, R.; García, A. E. Influence of Water-Protein Hydrogen Bonding on the Stability of Trp-Cage Miniprotein. A Comparison between the TIP3P and TIP4P-Ew Water Models. *Phys. Chem. Chem. Phys.* **2011**, *13* (44), 19840−19847.
(91) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin from ^{15}N NMR Relaxation. *J. Am. Chem. Soc.* **1995**, *117* (50), 12562−12566.
(92) Polyakov, K. M.; Lebedev, A. A.; Okorokov, A. L.; Panov, K. I.; Schulga, A. A.; Pavlovsky, A. G.; Karpeisky, M. Y.; Dodson, G. G. The Structure of Substrate-Free Microbial Ribonuclease Binase and of Its Complexes with 3′GMP and Sulfate Ions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58* (5), 744−750.
(93) Basconi, J. E.; Shirts, M. R. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2887−2899.
(94) Zondlo, S. C.; Lee, A. E.; Zondlo, N. J. Determinants of Specificity of MDM2 for the Activation Domains of p53 and p65: Proline27 Disrupts the MDM2-Binding Motif of p53. *Biochemistry* **2006**, *45* (39), 11945−11957.
(95) Richards, F. M.; Vithayathil, P. J. The Preparation of Subtilisin-Modified Ribonuclease and the Separation of the Peptide and Protein Components. *J. Biol. Chem.* **1959**, *234* (6), 1459−1465.
(96) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Cα Geometry: Φ, ψ and Cβ Deviation. *Proteins: Struct., Funct., Genet.* **2003**, *50* (3), 437−450.
(97) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10* (11), 5113−5124.
(98) Hansmann, U. H. E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(99) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Biopolymers* **2001**, *60* (2001), 96−123.
(100) Suárez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.; Zuckerman, D. M. Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. *J. Chem. Theory Comput.* **2014**, *10* (7), 2658−2667.
(101) Dickson, A.; Brooks, C. L. WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. *J. Phys. Chem. B* **2014**, *118* (13), 3532−3542.
(102) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys. J.* **1996**, *70* (1), 97−110.
(103) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400−1415.
(104) Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2012**, *8* (10), 3895−3901.
(105) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A.

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947

OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2015**, *12*, 281−296.
(106) Alexandrescu, A. T.; Rathgeb-Szabo, K.; Rumpel, K.; Jahnke, W.; Schulthess, T.; Kammerer, R. A. 15N Backbone Dynamics of the S-Peptide from Ribonuclease A in Its Free and S-Protein Bound Forms: Toward a Site-Specific Analysis of Entropy Changes upon Folding. *Protein Sci.* **1998**, *7*, 389−402.
(107) Jaravine, V. A.; Alexandrescu, A. T.; Grzesiek, S. Observation of the Closing of Individual Hydrogen Bonds during TFE-Induced Helix Formation in a Peptide. *Protein Sci.* **2001**, *10*, 943−950.
(108) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4* (8), 1−17.
(109) Wang, T.; Cai, S.; Zuiderweg, E. R. P. Temperature Dependence of Anisotropic Protein Backbone Dynamics. *J. Am. Chem. Soc.* **2003**, *125* (28), 8639−8643.

DOI: 10.1021/acs.jctc.6b00567
J. Chem. Theory Comput. 2016, 12, 3926−3947