JCTC Journal of Chemical Theory and Computation

Article

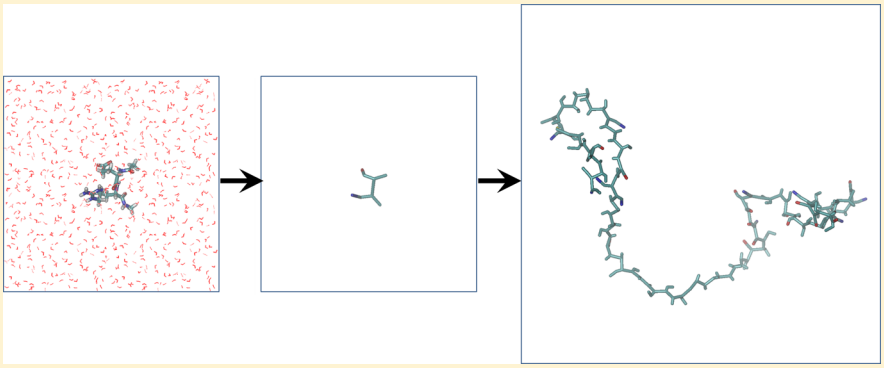
pubs.acs.org/JCTC

Parametrization of Backbone Flexibility in a Coarse-Grained Force Field for Proteins (COFFDROP) Derived from All-Atom Explicit-Solvent Molecular Dynamics Simulations of All Possible Two-Residue Peptides

Tamara Frembgen-Kesner, Casey T. Andrews, Shuxiang Li, Nguyet Anh Ngo, Scott A. Shubert, Aakash Jain, Oluwatoni J. Olayiwola, Mitch R. Weishaar, and Adrian H. Elcock*

Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, United States

Supporting Information



ABSTRACT: Recently, we reported the parametrization of a set of coarse-grained (CG) nonbonded potential functions, derived from all-atom explicit-solvent molecular dynamics (MD) simulations of amino acid pairs and designed for use in (implicit-solvent) Brownian dynamics (BD) simulations of proteins; this force field was named COFFDROP (COarse-grained Force Field for Dynamic Representations Of Proteins). Here, we describe the extension of COFFDROP to include bonded backbone terms derived from fitting to results of explicit-solvent MD simulations of all possible two-residue peptides containing the 20 standard amino acids, with histidine modeled in both its protonated and neutral forms. The iterative Boltzmann inversion (IBI) method was used to optimize new CG potential functions for backbone-related terms by attempting to reproduce angle, dihedral, and distance probability distributions generated by the MD simulations. In a simple test of the transferability of the extended force field, the angle, dihedral, and distance probability distributions obtained from BD simulations of 56 three-residue peptides were compared to results from corresponding explicit-solvent MD simulations. In a more challenging test of the COFFDROP force field, it was used to simulate eight intrinsically disordered proteins and was shown to quite accurately reproduce the experimental hydrodynamic radii (R_{hydro}), provided that the favorable nonbonded interactions of the force field were uniformly scaled downward in magnitude. Overall, the results indicate that the COFFDROP force field is likely to find use in modeling the conformational behavior of intrinsically disordered proteins and multidomain proteins connected by flexible linkers.

INTRODUCTION

Despite rapid advances in computer software and hardware, the computational demands of current large-scale biomolecular simulations routinely exceed computational capabilities,[1] and there is, therefore, continuing interest in the development of more rapid coarse-grained (CG) simulation models.[2−6] Although there are many methods available for developing CG potential functions, one common solution is to parametrize the CG model to reproduce data obtained from all-atom, explicit solvent MD simulations. This approach has been applied to a range of biomolecules, including DNA,[7−10] lipids,[11−18] carbohydrates,[19−24] and peptides and proteins.[25−32]

In the case of proteins, in particular, a number of studies have attempted to use explicit-solvent MD simulations to derive potential functions; to our knowledge, however, at the time of writing only two such studies appear to have done so by systematically simulating all possible pairwise interactions of two amino acids in order to derive generic, broadly applicable nonbonded potential functions.[26,32]

In the first of these previous studies, the pioneering work of the Betancourt group,[26] the GROMOS G43a1[33] force field and SPC[34] water models were used to develop potential functions for a one-bead-per-residue CG force field by direct Boltzmann inversion of radial distribution functions obtained from MD. In the second such study, the Amber ff99sb-ildn[35,36] force field

Received: January 15, 2015

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

was used together with the TIP4P-Ew[37] water model to develop potential functions for a multibead-per-residue force field via the iterative Boltzmann inversion (IBI) method;[38,39] the resulting set of CG nonbonded potential functions was named COFFDROP (COarse-grained Force Field for Dynamic Representations Of Proteins).[32] COFFDROP's (pairwise) nonbonded potential functions were then tested by comparison to additional explicit-solvent MD simulations that modeled multiple solute molecules at concentrations up to 300 mg/mL. Brownian dynamics (BD) simulations using COFFDROP were shown to provide excellent reproduction of the solute−solute interaction thermodynamics of pure amino acid systems, albeit at the expense of providing a limited ability to reproduce the fine structural details of the interactions.

Since the initial iteration of the COFFDROP force field was limited to describing nonbonded interactions, we report here the expansion of COFFDROP to include backbone potential functions. This study takes as its starting point the results of extensive explicit-solvent MD simulations of all combinations of two-residue peptides (441 systems in all) that we reported recently.[40] Here, we describe the derivation and testing of bonded potential functions for backbone pseudoatoms and the additional nonbonded potential functions required to describe the conformational energetics of peptides and proteins. Tests of the extended COFFDROP force field against further MD simulations and against experimental data on the hydrodynamic radii of intrinsically disordered proteins indicate that the resulting force field is likely to be useful for modeling proteins in a variety of scenarios.

METHODS

Systems Simulated Using Explicit-Solvent Molecular Dynamics. Most of the MD simulation data used to derive the potential functions reported here were described previously:[40] that paper described the results of MD simulations of all of the 441 possible two-residue peptides that can be constructed from the 20 standard amino acids with histidine modeled in both its protonated and unprotonated forms. In each of those simulations, the peptides were capped with acetyl (Ace) and *N*-methyl (Nme) groups at the N- and C-termini, respectively. In order to model proteins in general, however, it is also necessary to develop parameters for the uncapped, charged termini. To provide the raw simulation data required to derive bonded parameters for uncapped termini, we performed here two additional sets of MD simulations of each of the following amino acids: Ala, Gln, Gly, Lys, Pro, and Val. These amino acids were chosen because they provide representatives of each of the CG mapping schemes used in COFFDROP to represent amino acid side chains (see later): the Cβ pseudoatom of Gln, for example, maps to its parent all-atom model in exactly the same way as the Cβ pseudoatoms of Cys, Glu, Met, and Ser (Figure S1 and Table S1, Supporting Information). In one set of simulations, the N-terminus was uncapped (i.e., modeled as NH_3^+) and the C-terminus was capped with the Nme group; in the other set, the C-terminus was uncapped (modeled as COO^-), and the N-terminus was capped with the Ace group. The derivation of bonded parameters from these simulations is described in a later section.

To provide the raw MD simulation data needed to describe nonbonded interactions of the uncapped terminal groups with the amino acids, we used relatively long, unbiased explicit-solvent MD simulations to directly simulate their associations; we have previously used this "brute force" approach to measure the association thermodynamics of amino acids,[41−43] alkanes,[44] and nucleic acids.[45] Here, a total of 45 1 μs MD simulations were performed: (a) simulations of C-terminal capped Gly, i.e., $NH_3^+CH_2COCH_3$, with all 21 fully capped amino acids were used to derive parameters for the NH_3^+ group interacting with every amino acid; (b) simulations of N-terminal capped Gly, i.e., $CH_3NHCH_2COO^-$, with all 21 fully capped amino acids were used to derive parameters for COO^-, amino acid interactions; (c) a simulation of $NH_3^+CH_2COCH_3$ with $CH_3NHCH_2COO^-$ was used to obtain nonbonded parameters for the NH_3^+, COO^- interaction; (d) a simulation of two $NH_3^+CH_2COCH_3$ molecules was used to obtain nonbonded parameters for the NH_3^+, NH_3^+ interaction; and (e) a simulation of two $CH_3NHCH_2COO^-$ molecules was used to obtain nonbonded parameters for the COO^-, COO^- interaction.

In addition to the MD simulations used to derive the force field, we also performed MD simulations in order to test the force field's ability to predict the conformational behavior of more complicated systems. To this end, additional simulations were performed on a number of three-residue peptide systems. Of the 9261 possible three-residue peptides that can be constructed from all combinations of 21 amino acid types, a subset of 56 representative peptide systems were selected. Uncharged systems included AAA (using the one-letter codes for amino acids) as well as all of the combinations AAX, AXA, AXX, XAA, XAX, XXA, and XXX, with X in each case being Leu, Met, Phe, or Ser; these amino acids were selected in order to provide a sampling of systems containing side chains with aliphatic, aromatic, and polar characteristics. In addition to those 29 systems, 27 charged peptide systems were simulated that included all possible combinations of the three amino acids Gln, Glu, and Lys (i.e., EEE, EEK, EEQ, etc.).

MD Simulations. The MD simulations of the two-residue peptides used to derive bonded backbone and intramolecular nonbonded interactions have been described elsewhere.[40] The protocol used for simulating the additional systems outlined above was identical to that used previously with the sole exception of the length of the production periods of the simulations (see below). All MD simulations were performed using the GROMACS software package version 4.5.1[46,47] and the Amber ff99SB-ildn-NMR force field.[35,36,48] The peptides were placed in a periodic 35 × 35 × 35 Å^3 box and solvated with ∼1400 TIP4P-Ew[37] water molecules. Energy minimization of the systems was performed using the method of steepest descent for 1,000 steps. The systems were then heated in 50 K increments to 298 K over a period of 350 ps and allowed to equilibrate for a further 1 ns. Following this point, production MD simulations were performed for 300 ns in the case of the individual two-residue peptides, 1 μs for the interactions of pairs of amino acids, and 1 μs for individual three-residue peptides. All production simulations were conducted in the NPT ensemble; temperature (298.15 K) and pressure (1 atm) were maintained using the Nosé−Hoover thermostat[49,50] and the Parrinello−Rahman barostat,[51] respectively. Short-range nonbonded interactions were cut off at 10 Å, and the Particle Mesh Ewald (PME) method[52] was used to calculate all long-range electrostatic interactions. Covalent bonds were constrained to their equilibrium lengths using the LINCS algorithm,[53] allowing a 2.5 fs time step to be used. Solute coordinates were saved at intervals of 0.1 ps giving totals of 3 million or 10 million snapshots for production lengths of 300 ns and 1 μs, respectively.

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Derivation of COFFDROP Bonded and Nonbonded Potential Functions. As in our previous work,[32] the MD simulation data were used to generate target distributions of pseudoangles, pseudodihedrals, intramolecular nonbonded distances, and radial distribution functions, that were then used to guide the generation of CG potential functions. To this end, all solute snapshots sampled from each MD simulation were converted into their CG equivalents following the COFFDROP CG mapping scheme outlined in Figure S1 and Table S1 (Supporting Information). In this scheme, backbone atoms are represented by a single bead, or pseudoatom, placed at the Cα position of each residue, and side chains are modeled by between one (e.g., Ala) and three pseudoatoms (Trp only). For modeling capped peptide systems, additional pseudoatoms are placed at the methyl carbon positions of the acetyl (Ace) and *N*-methyl (Nme) capping groups; for modeling uncapped peptides, additional pseudoatoms are placed at the position of the N-terminal nitrogen and at the geometric center of the two C-terminal carboxylate oxygens (Figure S1, Supporting Information). After mapping each two-residue peptide MD trajectory into its CG equivalent, all pseudobond lengths were individually averaged, all probability distributions of pseudoangles and pseudodihedrals were binned in 5° intervals, and distance distributions for each pair of pseudoatoms involved in nonbonded interactions were binned in 0.1 Å intervals from 0 to 15 Å.

Since the purpose of the present work is to extend COFFDROP so that it is capable of describing backbone conformational preferences in peptide systems, all previously derived potential functions[32] were retained and left unaltered: this included all intraresidue pseudoangles, pseudodihedrals, and improper dihedrals, inter-residue nonbonded interactions, and, for Trp only, two intraresidue nonbonded interactions. The additional CG potential functions required to model polypeptide systems, i.e., bonded backbone potential functions and a number of nonbonded potential functions describing interactions of pseudoatoms on adjacent residues, were derived using the MD data on two-residue peptides as follows.

For reasons outlined in our earlier work,[32] we chose to model all pseudobonds using simple harmonic potential functions:

$$\varepsilon^{\mathrm{CG}} = K^{\mathrm{bond}}(x - x_0)^2 \tag{1}$$

where $\varepsilon^{\mathrm{CG}}$ is the energy of the bond, K^{bond} is the spring constant, x is the current length of the pseudobond in the CG simulation, and x_0 is its equilibrium length (obtained directly from analysis of the MD data). As in our previous work,[32] K^{bond} was set to 200 kcal/mol/Å^2 in order to allow a comparatively long 50 fs time step to be used in the BD simulations that employ the COFFDROP potential functions.

All other bonded and nonbonded potential functions were derived using the IBI method.[38,39] For all of the bonded and intramolecular nonbonded terms that were not previously parametrized in the original COFFDROP publication, initial guesses for the CG potential functions were generated by Boltzmann inversion of the MD probability distributions:

$$\varepsilon^{\mathrm{CG}}(\xi) = -RT\ln\{\mathrm{prob}^{\mathrm{MD}}(\xi)\} \tag{2}$$

where $\varepsilon^{\mathrm{CG}}(\xi)$ is a particular angle, dihedral, or intramolecular nonbonded potential function, R is the gas constant, T is the temperature in Kelvin, and $\mathrm{prob}^{\mathrm{MD}}(\xi)$ is the corresponding target probability distribution obtained from MD. These initial potential functions were combined with the previously derived potential functions and used to perform a 5 μs BD simulation. Probability distributions for all pseudoangles, pseudodihedrals, and intramolecular nonbonded interactions sampled during the BD simulation were then calculated and compared to their corresponding "target" distributions. The deviations of the distributions sampled during BD from the target MD distributions were then used to modify the potential functions according to

$$\varepsilon^{\mathrm{CG}}_{j+1}(\xi) = \varepsilon^{\mathrm{CG}}_{j}(\xi) + RT\ln\{\mathrm{prob}^{\mathrm{BD}}(\xi)/\mathrm{prob}^{\mathrm{MD}}(\xi)\} \times \alpha \tag{3}$$

where j is the current iteration number of the IBI procedure, $\mathrm{prob}^{\mathrm{BD}}(\xi)$ is the probability distribution sampled from BD, and α is a scaling factor employed to allow the process to converge more smoothly: a factor of 0.25 was used for all pseudoangles, pseudodihedrals, and intramolecular nonbonded terms.

The IBI procedure was conducted for 100 iterations for each amino acid system. Following each iteration, the absolute error between the distributions computed during BD and the target distributions obtained from MD was calculated using

$$\mathrm{Err}(\xi) = \sum_{k=1}^{N} |\mathrm{prob}^{\mathrm{MD}}_{k}(\xi) - \mathrm{prob}^{\mathrm{BD}}_{k}(\xi)| \tag{4}$$

where N is the number of potential functions being optimized. The sum extends over each bin of the distribution (36, 72, and 150 bins for pseudoangles, pseudodihedrals, and intramolecular nonbonded terms, respectively). As in our previous work,[32] the iteration that produced the lowest overall error in the pseudodihedral distributions was selected for inclusion in the final force field.

For the derivation of CG potential functions for the pseudoatoms used to represent the NH_3^+ and COO^- groups in uncapped peptides, the parametrization procedure followed was the same as that reported previously:[32] bonded and intramolecular nonbonded potential functions were derived from MD simulations of single amino acids, while intermolecular nonbonded potential functions were derived from MD simulations of amino acid pairs. Application of the IBI method to simulation data for single amino acids with uncapped charged groups allowed us to optimize CG potential functions describing each of the following pseudoangles (Ace−Cα−COO^-, Cβ−Cα−COO^-, NH_3^+−Cα−Nme, and NH_3^+−Cα−Cβ) and improper pseudodihedrals (Ace−COO^-−Cα−Cβ, COO^-−Ace−Cα−Cβ, NH_3^+−Nme−Cα−Cβ, and Nme−NH_3^+−Cα−Cβ). Since the mapping scheme that we use to convert all-atom models to their CG equivalents uses identical atoms to describe the Cβ pseudoatom of Val, Ile, and Thr, we assumed that the bonded CG potential functions derived for uncapped Val could be applied directly to Ile and Thr. For similar reasons, we assumed that the bonded CG potential functions derived for Gln could be applied to Cys, Glu, Met, and Ser, those derived for Lys could be applied to Arg, and those functions derived for Ala could be applied to Asp, Phe, His, protonated His (Hip), Leu, Asn, Trp, and Tyr (Figure S1 and Table S1, Supporting Information).

We note that adopting this strategy does not provide us with a complete parametrization of the uncapped terminal residues as certain side chain pseudodihedral functions involving the charged termini, for example, NH_3^+−Cα−Cβ−Cγ in Trp, remain unparameterized. In the few simulations reported here that used these unparameterized pseudodihedral terms

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(simulations of intrinsically disordered proteins (IDPs); see below), they were modeled with simple cosine potential functions described in our previous publication:[32]

$$E(\phi) = V_1[1 + \cos(\varphi - \varphi_1)] + V_3[1 + \cos(3\varphi - \varphi_3)] \quad (5)$$

where V_1 is 0.5 kcal/mol/rad, V_3 is 0.25 kcal/mol/rad, ϕ is the value of the dihedral, and φ_1 and φ_3 are the phase angles defining the position of the energy maxima of the cosine term (defined here using a representative initial model of each IDP). A more complete parametrization of the uncapped, charged termini will be explored in a future iteration of COFFDROP.

Intermolecular nonbonded potential functions involving the NH_3^+ and COO^- groups were derived from 45 MD simulations of amino acid pairs (see above). As in our previous work,[32] the initial CG nonbonded potential functions were assigned a $1/r_{ij}^{12}$ potential, where r_{ij} is the distance between two pseudoatoms i and j. Thereafter, the potential function of each intermolecular nonbonded term was iteratively optimized to reproduce its radial distribution function, $g(r)$, using an equation analogous to eq 3; to limit oscillations in the IBI procedure, a damping factor, α, of 0.05 was used. Each target $g(r)$ was calculated as the ratio of the distance distribution obtained from the MD simulation to the corresponding distribution obtained from 100 million random placements of the same pair of amino acids in the simulation box. As in our previous work,[32] the adjusted nonbonded potential functions were each smoothed twice using a Savitsky–Golay[54] procedure prior to performing the subsequent iteration of the optimization process.

Brownian Dynamics Simulations of Peptides. All BD simulations were performed using the algorithm of Ermak and McCammon[55] with in-house software. Settings for the simulation of all peptide systems replicated those of the corresponding MD simulations as closely as possible. A 35 × 35 × 35 Å^3 box was used for all simulations, and a grid-based Ewald method[56] was used to calculate the long-range electrostatic interactions in simulations that contained two or more charged pseudoatoms; energies and forces calculated using this method were identical to those calculated using the smooth PME method implemented in GROMACS. For simulations of capped peptide systems, the only pseudoatoms carrying a nonzero partial charge were the pseudoatoms closest to the true position of the charged functional group in Arg, Lys, and protonated His side chains (to which a +1 charge is assigned) and Asp and Glu side chains (to which a −1 charge is assigned); for simulations of uncapped peptide systems, +1 and −1 charges were assigned to the pseudoatoms representing the NH_3^+ and COO^- groups, respectively. The dielectric constant for electrostatic calculations in all peptide simulations was set to 62.9, i.e., the literature value reported for the TIP4P-Ew water model[37] used in the MD simulations. All BD simulations of peptide systems were conducted for 5 μs using a time step of 50 fs. In order to ensure realistic translational and rotational diffusion of the solutes,[57] intramolecular hydrodynamic interactions were modeled at the Rotne–Prager–Yamakawa level of theory,[58,59] with correlated random displacements computed using a Cholesky decomposition of each intramolecular diffusion tensor; the diffusion tensors were updated every 20 ps. Hydrodynamic radii of 3.5 Å were assigned to all pseudoatoms, and a viscosity of 0.89 cP, corresponding to that of pure water at 298 K, was used throughout.

Intrinsically Disordered Proteins. COFFDROP's ability to model the conformational properties of protein systems was tested by modeling a number of IDPs. These were selected from a list of 32 peptides and proteins whose experimental hydrodynamic properties were summarized and analyzed recently.[60] Of these, we selected only those proteins with a sequence composition of less than 5% prolines: the current version of COFFDROP models all peptide bonds in the trans configuration and so is not expected to perform well at describing the conformational properties of proteins that may have significant numbers of peptide bonds in the cis configuration. We also selected only those proteins whose hydrodynamic properties have been measured at experimental pH values between 6.5 and 7.5; this decision was based on the fact that the parameters of the ionizable residues other than histidine have been derived at charge states most appropriate to pH 7.

Application of these selection criteria gave a reduced list of 10 proteins, from which the C-terminal domain of L9[61] was eliminated on the basis that it is disordered only at low temperature,[61] and from which the anticodon-binding domain of tyrosyl-tRNA synthetase[62] was also eliminated as its structure has been solved by NMR in the absence of binding partners (suggesting that it is not, therefore, intrinsically disordered).[63] In all, therefore, eight IDPs were selected for simulation; these were the amyloid $\beta_{(1-40)}$ peptide (A$\beta_{(1-40)}$),[64] suppressor of Mec1 lethality (Sml1),[65] *Lotus japonicas* IDP 1 (*Lj*IDP1),[66] prothymosin α (ProTα),[67] abscisic acid stress ripening 1 (ASR1),[68] the nucleoporin Nup116,[69] α-Synuclein (α-Syn),[70] and the cystic fibrosis transmembrane conductance regulator regulatory region (CFTR R).[71]

The hydrodynamic radius (R_{hydro}) of each IDP and the pH of the experiment used to measure R_{hydro} were verified from the original source material; in those cases where a translational diffusion coefficient was reported, this was converted to a hydrodynamic radius using the Stokes–Einstein relationship ($R_{hydro} = k_BT/6\pi\eta D_{trans}$, where k_B is Boltzmann's constant, T is the temperature in K, η is the viscosity of the solvent, and D_{trans} is the translational diffusion coefficient) with the appropriate temperature and viscosity. The amino acid sequence of each IDP was taken from the original source or, if absent, was taken from the UniProt Knowledge Base (www.uniprot.org).[72] Details of each simulated IDP are listed in Table S2 (Supporting Information).

Initial atomic models of each IDP were constructed by first generating random backbone conformations with the program RCG,[73] then adding side chains with the program SCWRL4.0.[74] A total of 100 different conformations were built for each protein; these were ordered by increasing radius of gyration, and 10 conformations were selected as starting points for BD simulation by taking every 10th structure starting with the fifth ranked structure and ending with the 95th ranked structure. For ionizable residues, charges were assigned based on average pK_a values of each type of side chain reported by Antosiewicz et al.[75] and the pH of the experiment; pK_a values of 2.7, 4.0, 12.0, 10.1, and 6.9 were assumed for Asp, Glu, Arg, Lys, and His, respectively. The same approach was used to assign charges to the pseudoatoms of the N and C termini, for which pK_a values of 8.2 and 2.7 were used, respectively.[75] In pursuing this approach, therefore, we are assuming that the nonelectrostatic components of the nonbonded interactions derived for ionizable residue types remain unaltered by changing pH and that electrostatic components can be obtained

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

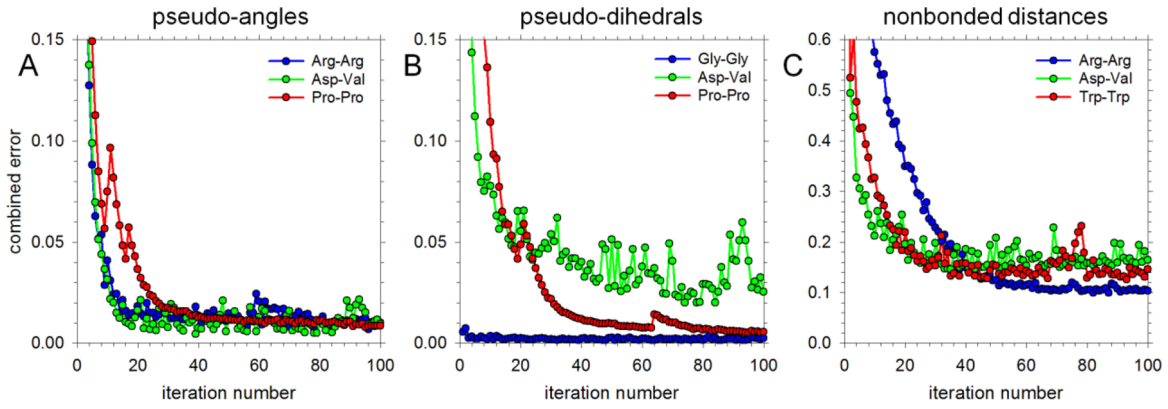


Figure 1. Optimization of bonded and nonbonded potential functions using the IBI method. (A) Combined absolute error in pseudoangle probability distributions of three representative systems versus iteration number. (B) Same as A but showing results for pseudodihedral probability distributions of selected systems. (C) Same as A but showing results for distance probability distributions of selected systems.

by simply scaling the charge to the average value expected using the Henderson−Hasselbalch equation. For histidines, we used the nonbonded parameters derived for the neutral (His) residue type in all simulations; use of the nonbonded parameters for the protonated form of the residue will be reserved for future work intended to explicitly model changes in protonation state.

BD simulations of IDPs were performed using a protocol very similar to that used in simulations of small peptides with the exception that all simulations were performed without periodic boundary conditions applied. Since all 8 IDPs were studied at a temperature within 3 K of 298 K, the temperature in all BD simulations was set to this temperature, and both the viscosity and the dielectric constant were set to their corresponding experimental values at that temperature (0.89 cP and 78.4, respectively). To explore the extent to which the simulated behavior might depend on the value assigned to the dielectric constant, we carried out additional BD simulations on the representative IDP ASR1 using the same dielectric constant (62.9) used in simulations of the peptides. We obtained identical estimates of R_{hydro} with the two dielectric constants: 27.0 ± 0.4 Å with a dielectric constant of 78.4 and 27.1 ± 0.5 Å with a dielectric constant of 62.9 (see below for details of how R_{hydro} was calculated).

The ionic strengths and pH values used in the BD simulations were identical to those used experimentally; the effects of the former were incorporated in a simple way by assuming Debye−Hückel screening of the Coulombic interactions, i.e., $E_{Debye-Hückel} = 332 q_i q_j \exp(-\kappa r_{ij})/78.4 r_{ij}$, where q_i and q_j are the pseudoatomic charges, κ is the Debye−Hückel screening parameter, related to the square root of the ionic strength, and r_{ij} is the distance between the charges. All long-range electrostatic interactions were modeled using a treecode method,[76] with the forces obtained from the treecode being updated every 20 ps: use of the treecode means that simulations can be performed in a straightforward way without imposing periodic boundary conditions and thereby allowing us to more easily mimic the very dilute conditions studied experimentally. Initial simulations of the IDPs indicated that results converged within 3 μs of production time; to be conservative, however, all simulations were run for 4 μs with snapshots collected every 10 ns for analysis.

In order to compare with the experiment, the hydrodynamics program HYDROPRO10[77] was used to compute R_{hydro} values for all snapshots of the IDPs. HYDROPRO10 is most commonly used to compute the hydrodynamic properties of atomic models of proteins and nucleic acids and, in particular, has been shown to produce R_{hydro} values in very good agreement with experimental values when a hydrodynamic radius of 2.9 Å is assigned to the "beads" that are used to represent the all-atom models.[77] To determine the best hydrodynamic radius to assign to HYDROPRO10's beads when applied to our CG models, we proceeded as follows. First, we used HYDROPRO10 with the currently recommended bead size of 2.9 Å to calculate the R_{hydro} values of all-atom models of 11 folded, globular proteins (see ref 57 and references therein) and 10 RCG-generated structures of each IDP. Next, we repeated the calculations using the CG versions of the proteins with a range of bead sizes (Figure S2 in Supporting Information) and selected the one that best reproduced the results obtained with the corresponding all-atom model. For folded proteins, an optimal bead radius of 3.44 Å was obtained; for IDPs, an optimal bead radius of 3.42 Å was found; these results were averaged to give a final bead radius of 3.43 Å.

As was found in our previous work,[32] the scaling of COFFDROP's nonbonded potential functions improved its ability to reproduce experimental data (see Results). To this end, simulations of all IDPs were performed using a range of nonbonded scaling factors from 0.5 to 1.0 in intervals of 0.1. The final averaged R_{hydro} values from simulations performed using these different scaling factors were fit to the four-parameter sigmoidal equation:

$$R_{hydro}(x) = c + a/\{1 + e^{-(x-x_0)/b}\} \quad (6)$$

where x is the scaling factor applied to the nonbonded potential functions, and a, b, c and x_0 are parameters. Solving for x gave us the scaling factor needed for each protein to reproduce the experimental R_{hydro} value. These individual scaling factors were averaged in order to obtain a single, transferable scaling factor that could be used in simulations of other proteins.

RESULTS

Derivation of Bonded and Nonbonded Potential Functions. The two-residue peptide MD simulation data on which much of this work is based were described and analyzed by Li et al.[40] The 3,000,000 snapshots obtained from each MD simulation were converted to their CG equivalents, and "target" pseudoangle, pseudodihedral, and distance distributions were

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

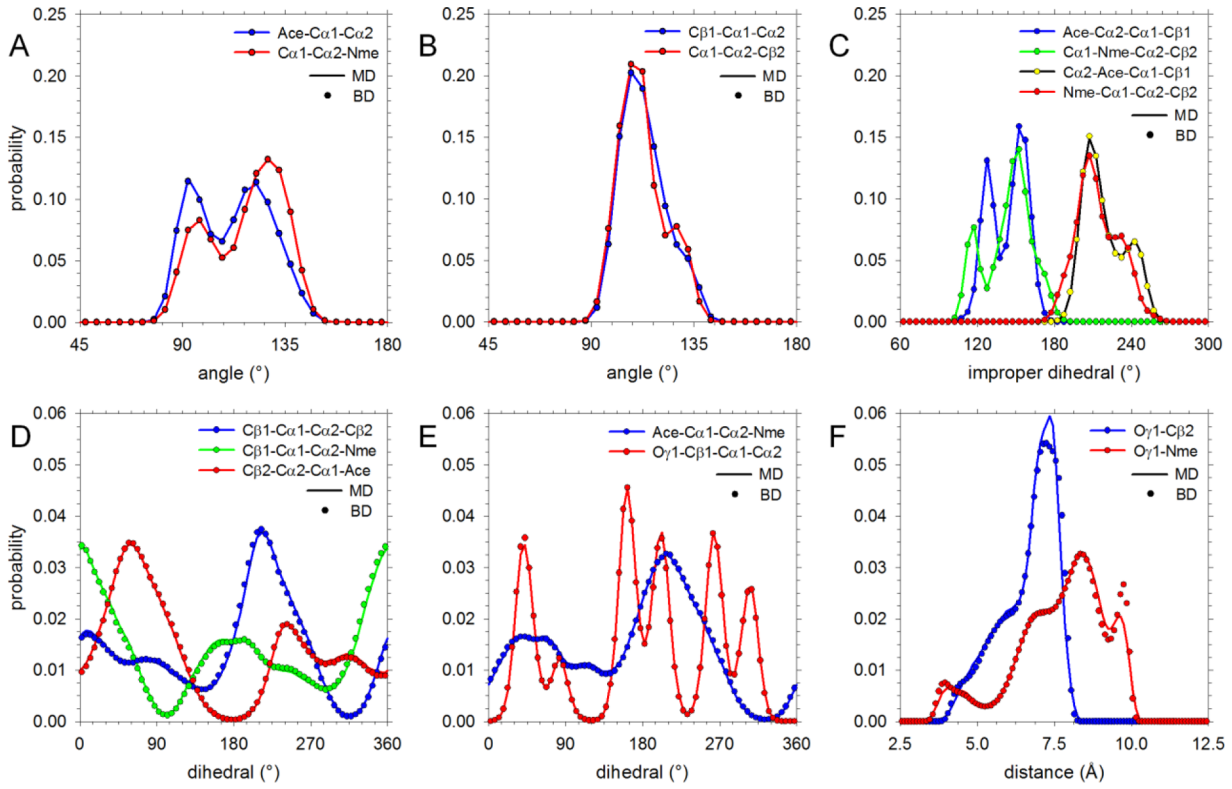


Figure 2. Comparison of probability distributions obtained from a BD simulation using optimized CG potential functions with target probability distributions from MD. Results shown are for the Asp-Val peptide; MD results are denoted by solid lines and BD results by filled circles. (A,B) Results for the four pseudoangles. (C) Results for improper dihedrals. (D,E) Results for pseudodihedrals. (F) Results for intramolecular nonbonded terms. In all cases, numbers assigned to pseudoatoms indicate the residue: Cα1 is the Cα pseudoatom of Asp; Cα2 is the Cα pseudoatom of Val.

calculated. The IBI method[38,39] was then used to simultaneously optimize a variety of bonded and nonbonded CG potential functions in order to reproduce these target distributions (see Methods). The total absolute differences between the distributions sampled during each of the 100 iterations of BD simulation and the target MD distributions are plotted in Figure 1 for select systems: Asp-Val, for example, is shown because its flexibility is closest to the mean of the 441 peptides studied by Li et al.;[40] Gly-Gly and Pro-Pro are shown because they are, respectively, the most flexible and rigid of the systems simulated.

In all cases, the decreases in error are most substantial during the first ~20 iterations of the IBI procedure, with a general leveling-off at ~40 iterations. The overall errors of the pseudoangle probability distributions are low (e.g., Figure 1A), with little to no noise, while those of the pseudodihedral probability distributions are noisier (e.g., Figure 1B), particularly for the Asp-Val system, which has 9 dihedral functions to simultaneously optimize. The errors in the intramolecular nonbonded distance distributions (Figure 1C) appear to be generally higher than those of the pseudoangle and pseudodihedral terms, but this is likely to be at least partly due to the combined error calculation for these terms involving a greater number of bins (see Methods).

Importantly, however, the probability distributions obtained from the BD simulation using the optimized potential functions are in all cases a near-perfect fit to the target MD probability distributions. Figure 2 compares the MD-derived probability distributions (lines) with the corresponding BD-derived probability distributions (circles) for a variety of terms in the Asp-Val system: in all panels, good agreement would be indicated by symbols lying on or close to the corresponding lines. Agreement between MD and BD is especially good for the pseudoangle and pseudodihedral terms (Figures 2A–E), including the four improper dihedral functions that were incorporated, as in our previous work,[32] in order to prevent the amino acids from interconverting between L- and D-like configurations (Figure 2C). The distance probability distributions for the intramolecular nonbonded interactions (Figure 2F) are in less good agreement but still capture the essential features of the corresponding MD distributions. Similar results are shown for the 14 different intramolecular nonbonded functions of the Trp-Trp peptide in Figure S3 (Supporting Information).

A combined comparison of all pseudoangle probability values sampled during BD with those sampled during MD is shown in Figure 3A for the 1724 terms whose potential functions have been optimized by IBI here; a corresponding plot for the pseudodihedral probability values (3951 terms) is shown in Figure 3B. The agreement in both cases is quite good, indicating that the IBI procedure succeeds in optimizing the bonded potential functions that it has been asked to adjust. A comparison of the pseudoangle probability values sampled during BD with those sampled during MD is shown in Figure 3C for the 1468 terms whose potential functions were *not* optimized by IBI here but which were instead retained from our previous work on single amino acids;[32] again, a corresponding plot for the pseudodihedral functions (628 terms) is shown in Figure 3D. The fact that the agreement remains reasonable suggests that potential functions that were previously derived from MD data on single amino acids[32] remain useable for two-residue peptides, but the scatter in the data also illustrates the

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

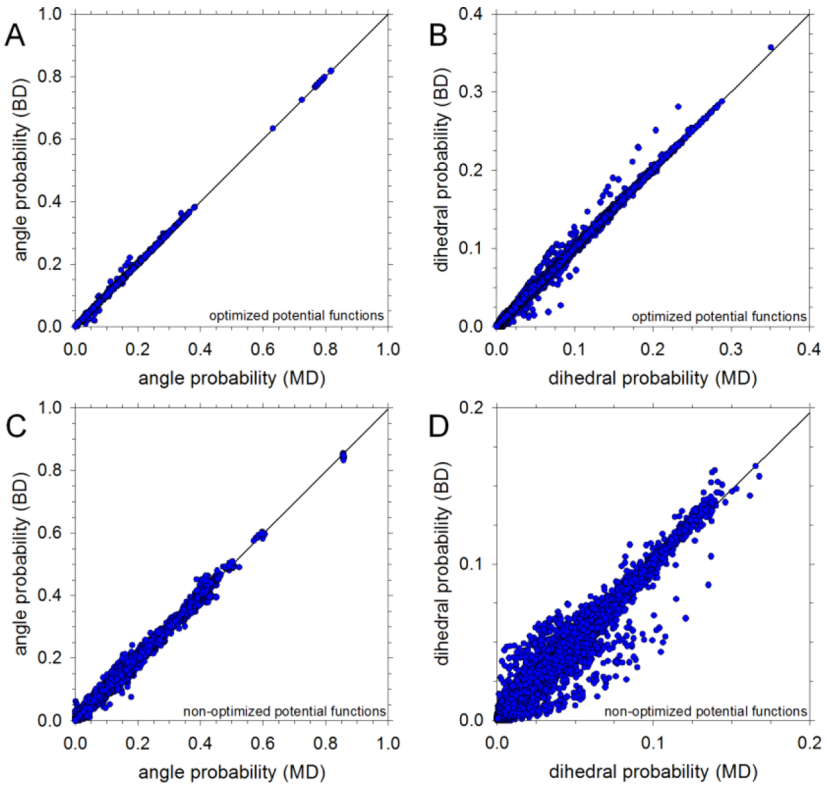


Figure 3. Comparison of all probability values obtained from BD simulations using optimized CG potential functions with target probability values obtained from MD for all 441 two-residue systems. (A) Comparison of pseudoangle probabilities for all optimized potential functions. (B) Same as A but showing results for pseudodihedrals. (C) Same as A but showing results for all nonoptimized potential functions, i.e., functions previously derived by Andrews and Elcock.[32] (D) Same as C but showing results for pseudodihedrals.

potential limits of the transferability of COFFDROP's bonded potential functions.

The significant iteration-dependent decreases in the errors that are shown in Figure 1 are strong arguments for using the IBI method to optimize the potential functions obtained initially from Boltzmann inversion (BI) of the target probability distributions. This is further illustrated in Figure 4, where the initial potential functions (lines), obtained by BI of the observed distributions, are compared with the optimized potential functions (circles) obtained by use of IBI for the representative Asp-Val system. Notably, while there is generally reasonable agreement for the pseudoangles and for the improper dihedrals (Figure 4A, B, and C), there are drastic differences for some of the pseudodihedral functions (Figure 4D and E). In addition, there is a reversal of the global and local minima for the Cα1−Cα2−Nme pseudoangle and for the Oγ1−Nme intramolecular nonbonded interaction (Figure 4A and F respectively).

To explore the robustness of the IBI method, we asked to what extent the final, optimized potential functions would change if different initial potential functions were assigned at the start of the IBI procedure. To this end, we repeated the IBI procedure for the Asp-Val system using as initial guesses the optimized potential functions obtained for the Leu-Cys system. Figure 5A compares the potential functions for a representative pseudoangle energy function (Cα1−Cα2−Cβ2): in this case, the initially assigned "wrong" energy function (green line) is clearly different from the potential function optimized using the standard procedure (blue line). After IBI, however, the reoptimized energy function (red upward triangles) is essentially identical to the original COFFDROP potential function. Figure 5B plots the probability distributions corresponding to the "wrong" and the reoptimized energy functions shown in Figure 5A, and compares them with the target MD probability distribution; as anticipated, potential functions reoptimized from even quite poor initial guesses are capable of reproducing the target simulation data. Figure 5C and D shows similar comparisons for a representative pseudodihedral energy function (Oγ1−Cβ1−Cα1−Cα2); again, these indicate that, within reason, the energy functions used to initialize the IBI procedure need not be a major determinant of the final potential functions.

As a test of the robustness of a different aspect of our strategy, we also investigated the extent to which the use of a different MD data set replicate would cause differences in the optimized potential functions. To this end, we performed three independent MD simulations of the Asp-Val system and used each as the basis of a separate IBI procedure; the three resulting sets of potential functions, shown in Figure S4 (Supporting Information), are very similar to each other, indicating that the 300 ns trajectories used as the basis of the parameters derived here are likely to be sufficiently converged to provide a reasonable reflection of the underlying MD force field.

Testing COFFDROP: Three-Residue Peptides. With the parameters of the force field derived from simulations of two-residue peptides, the most straightforward initial test of the expanded force field is to examine how well it predicts the conformational behavior of longer peptides. We make a number of assumptions in applying COFFDROP to such systems. First, we assume that pseudoangle and pseudodihedral functions necessary to describe the backbone conformational behavior of 3 or more consecutive residues can be obtained

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

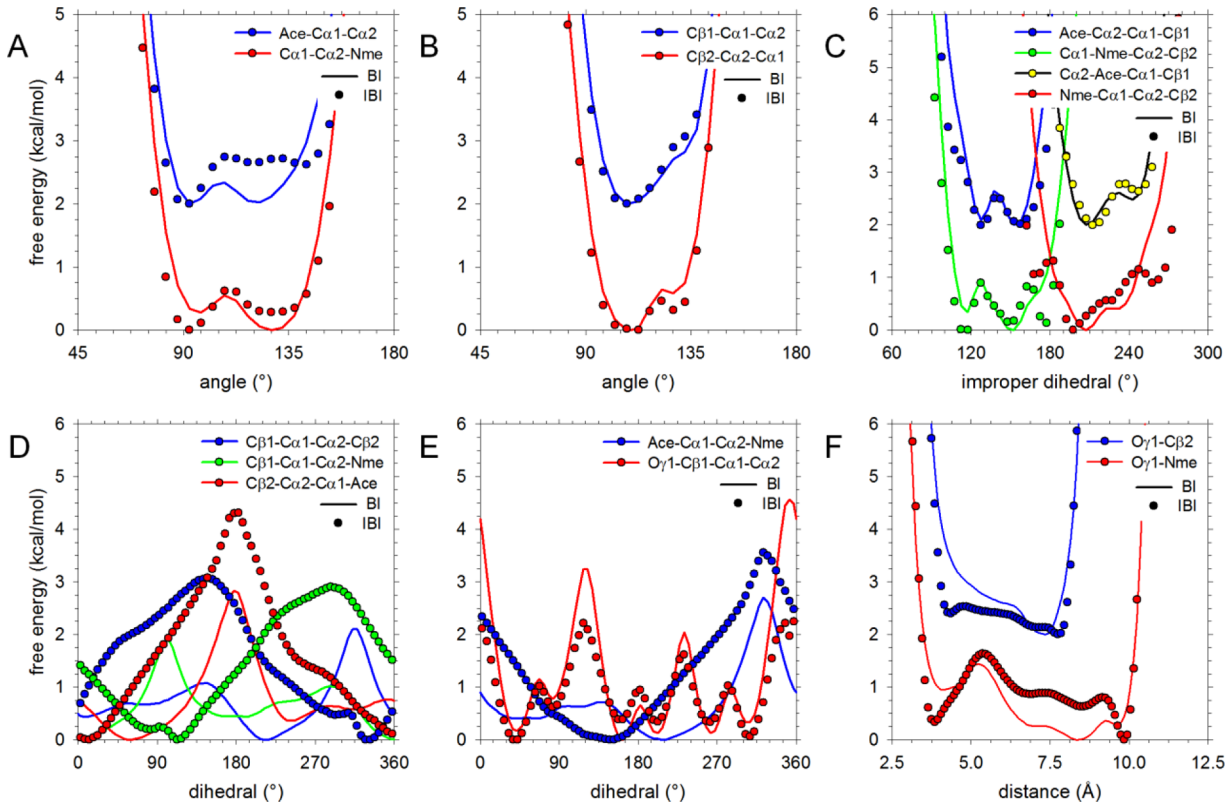


Figure 4. Comparison of optimized potential functions obtained from IBI with initial potential functions assigned by Boltzmann inversion (BI). Results shown are for Asp-Val; IBI results are denoted by filled circles and BI results by solid lines. (A,B) Results for pseudoangles; certain functions have been shifted upward for clarity. (C) Results for improper dihedrals; certain functions have been shifted upward for clarity. (D,E) Results for pseudodihedrals. (F) Results for nonbonded interactions; certain functions have been shifted upward for clarity.

directly from those derived here from simulations of capped, two-residue peptides. This means, for example, that the Cα−Cα−Cα−Cα pseudodihedral function needed to describe sequences of the form Xxx−Asp−Val−Yyy (where Xxx and Yyy are any amino acid) is taken from that of the Ace−Cα−Cα−Nme pseudodihedral function derived from a simulation of (capped) Asp−Val. Second, we assume that *intra*molecular nonbonded potential functions involving pseudoatoms on two residues separated by one or more residues can be described using nonbonded potential functions that were derived from *inter*molecular interactions of two amino acids. This means that the nonbonded interactions derived originally from simulations of a (capped) Asp + (capped) Val system[32] are used to describe nonbonded interactions of side chain atoms in sequences, for example, of the form Asp−Xxx−Val, Asp−Xxx−Yyy−Val, or Asp−Xxx−Yyy−Zzz-Val, etc.; the nonbonded interactions of side chain atoms in sequences of the form Asp−Val, however, come directly from those derived (here) from simulations of (capped) Asp−Val.

With those assumptions made, we have compared results from BD simulations of a set of 56 three-residue peptides directly against results from their MD counterparts (see Methods). Probability distributions were calculated and compared for a variety of pseudoangles, pseudodihedrals, and intramolecular nonbonded distances and the similarities of the MD and BD distributions were evaluated by calculation of correlation coefficients summarized in Table 1. As hoped, the correlation coefficients are generally very good. Figure 6A shows representative results for the pseudoangles of the Phe-Ala-Phe system, chosen because its correlation coefficients are closest to the median values obtained from the 56 systems tested. The agreement between MD (lines) and BD (symbols) is not exact, but is in each case good, with the positions of the peaks and shoulders in the distributions coinciding well. Figure 6B shows corresponding results for the pseudodihedral distributions of the Glu-Lys-Glu system, again chosen because it is representative; the positions and relative heights of the peaks are reproduced quite well. Finally, Figure 6C compares distance distributions for pseudoatom pairs involved in nonbonded interactions in the Lys-Lys-Glu system, with similarly good results being obtained.

While in most cases the correspondence between the behavior predicted by COFFDROP and the behavior observed in MD was very good, 3 out of the 56 systems (Gln-Glu-Gln, Glu-Gln-Lys, and Glu-Glu-Glu) produced some distributions that were in quite poor agreement. Figure S5 (Supporting Information) shows the single worst result that we obtained, for the dihedral functions of the Glu-Glu-Glu system. Although we were unable to determine common factors that would explain the poor results for these particular systems, we have been able to determine that much better results can be obtained using potential functions optimized specifically to reproduce the three-residue MD data. We explored this in two ways. In one approach ("full" optimization), we carried out a complete optimization of *all* potential functions, i.e., we retained none of the original potential functions from our earlier work[32] or derived from the two-residue simulations here. In a second approach ("partial" optimization), we optimized only those potential functions that were unique to the three-residue peptides and that, therefore, had not been explicitly para-

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

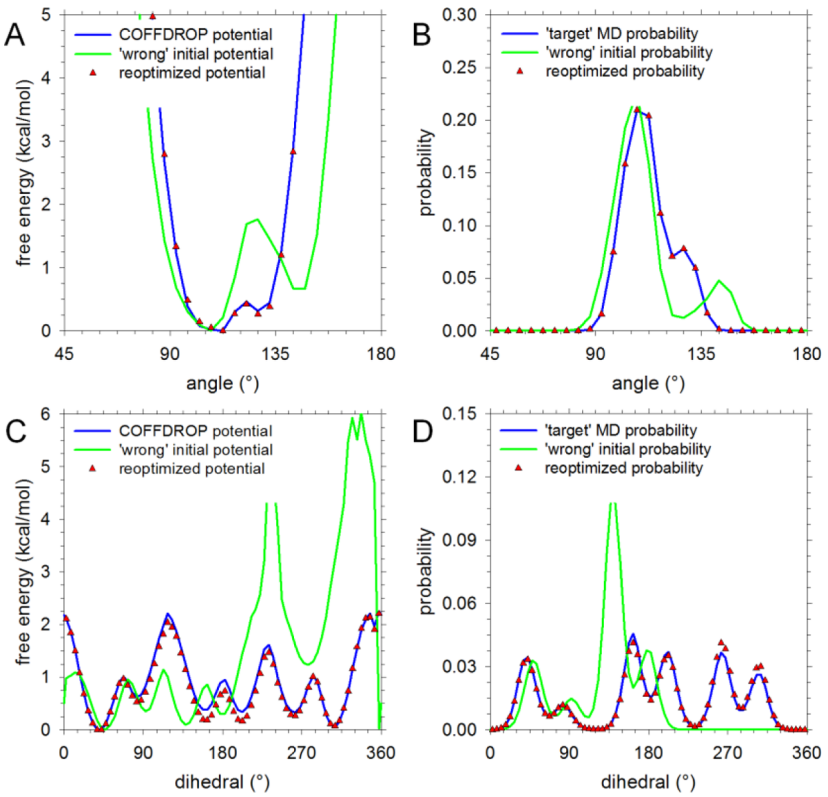


Figure 5. Comparison of results obtained from IBI starting with different initial potential functions. Results shown are for Asp-Val. (A) Representative pseudoangle energy function. (B) Corresponding probability distribution. (C) Representative pseudodihedral energy function. (D) Corresponding probability distribution.

Table 1. Correlation Coefficients between Probability Distributions Sampled from BD and MD for Three-Residue Peptides

	min.	max.	median
nonbonded interactions:			
Cβ1−Cβ2	0.908	0.999	0.984
Cβ2−Cβ3	0.909	0.998	0.984
Cβ1−Cβ3	0.529	0.996	0.985
Cγ1−Cβ2	0.860	0.999	0.988
Cγ1−Cβ3	0.715	0.994	0.977
Cγ2−Cβ1	0.797	0.997	0.986
Cγ2−Cβ3	0.865	0.998	0.990
Cγ3−Cβ1	0.681	0.998	0.974
Cγ3−Cβ2	0.892	0.997	0.987
Cγ1−Cγ2	0.932	0.999	0.989
Cγ2−Cγ3	0.933	0.998	0.990
Cγ1−Cγ3	0.783	0.996	0.982
pseudoangle terms:			
Ace−Cα1−Cα2	0.822	1.000	0.988
Cα1−Cα2−Cα3	0.792	0.999	0.981
Cα2−Cα3 − Nme	0.899	1.000	0.996
pseudodihedral terms:			
Ace−Cα1−Cα2−Cα3	0.686	0.998	0.964
Cα1−Cα2−Cα3−Nme	0.532	0.997	0.964
Cβ1−Cα1−Cα2−Cβ2	0.668	0.998	0.960
Cβ2−Cα2−Cα3−Cβ3	0.533	0.998	0.963

metrized before (e.g., the Cα-Cα-Cα angle potential function for Gln-Glu-Gln). Table S3 (Supporting Information) compares the correlation coefficients obtained between BD and MD using the COFFDROP parameters ("original") with those obtained when potential functions were optimized specifically to match the three-residue MD data. As expected, the results improve significantly. More interestingly, however, the results obtained with the "partial" optimization are very similar to those obtained with "full" optimization; this indicates that retaining potential functions previously derived from smaller systems and optimizing only those potential functions that are "new" could be a viable approach to parametrizing COFFDROP for larger or more complex peptide systems.

Testing COFFDROP: Intrinsically Disordered Proteins (IDPs). Finally, we tested COFFDROP on a number of protein systems, specifically 8 IDPs selected from a list of 32 reviewed and analyzed by Marsh and Forman-Kay[60] (see Methods and Table S2, Supporting Information). A preliminary analysis of snapshots sampled from BD simulations performed with COFFDROP produced hydrodynamic radii, R_{hydro}, that were too small relative to experimental values (data not shown). As in our earlier work,[32] therefore, we repeated the simulations of the 8 IDPs with all energetically favorable nonbonded interactions scaled by factors of between 0.5 and 1.0. For all 8 proteins, independent simulations were performed using 10 very different initial structures (see Methods); in all cases, simulations were run for sufficiently long periods of time that converged predictions of the R_{hydro} values were obtained (Figure S6, Supporting Information). The resulting average computed R_{hydro} values for each IDP are plotted as a function of the nonbonded force field scaling factor in Figure 7A and B. Interestingly, for all proteins, the data points fit very well to a four-parameter sigmoidal function (Methods): at high values of the scaling factor, the IDPs collapse, and their R_{hydro} values reach a lower limit, while at low values of the scaling factor the

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

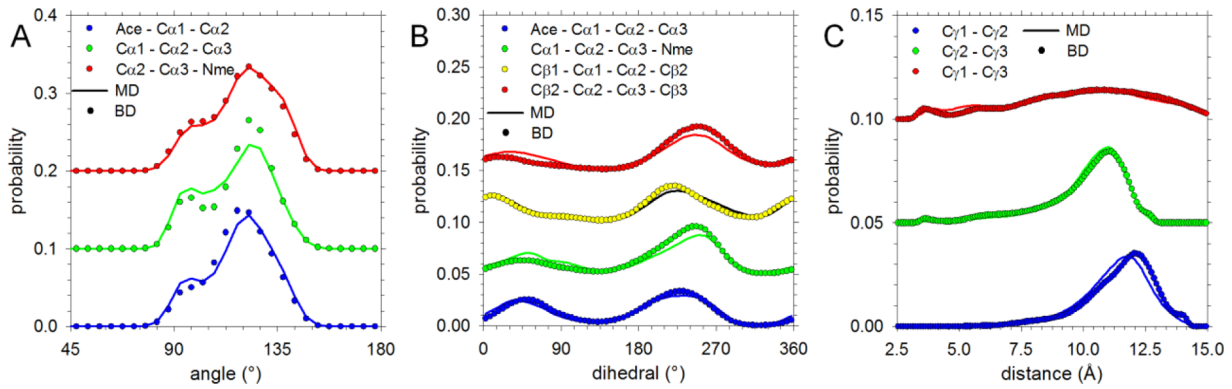


Figure 6. Comparison of probability distributions obtained from BD simulations of three-residue peptides with those obtained from MD. (A) Representative pseudoangle distributions for the Phe-Ala-Phe peptide; certain functions have been shifted upward for clarity. (B) Same as A but showing representative pseudodihedral distributions for the Glu-Lys-Glu peptide. (C) Same as A but showing representative distance distributions for the Lys-Lys-Glu peptide.

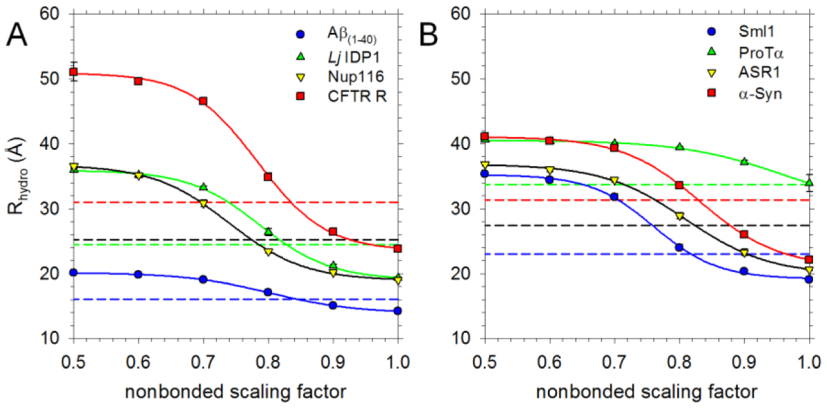


Figure 7. R_{hydro} values of IDPs obtained from BD simulations as a function of the nonbonded scaling factor. Results for the 8 IDPs are arbitrarily separated into A and B for clarity. Simulation results are plotted as filled circles with error bars (generally smaller than the symbols) indicating the standard deviation of 10 independent simulation results. Solid lines indicate fits of the data to the four-parameter sigmoidal function shown in eq 6; dashed horizontal lines indicate the corresponding experimental R_{hydro} values.

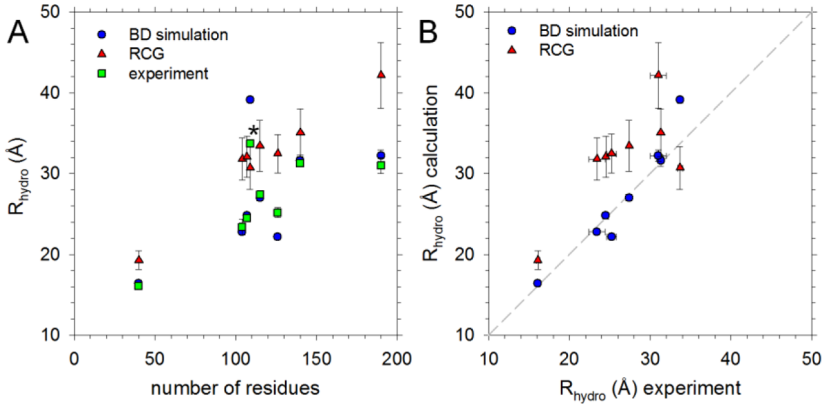


Figure 8. R_{hydro} values of IDPs compared with experimental values. (A) R_{hydro} values obtained from BD simulation using a nonbonded scaling factor of 0.825 (blue), RCG (red), and experimental values (green) plotted vs number of residues in the IDP; the asterisk indicates the BD simulation result for ProTα. (B) R_{hydro} values obtained from BD simulation using a nonbonded scaling factor of 0.825 (blue) and from RCG (red) plotted versus corresponding experimental values; the gray dashed line indicates perfect correlation.

IDPs adopt highly expanded states. For 6 of the IDPs, the best fits to the experimental R_{hydro} values occur with scaling factors of between 0.8 and 0.9; for one IDP (Nup116) the optimal scaling factor is ~0.75, while for the final IDP (ProTα) the optimal scaling factor is ~1.0. For reasons outlined in the Discussion, we discarded the fit for ProTα as an outlier and used the remaining scaling factors to obtain a compromise best-fit scaling factor of 0.825.

We then simulated all 8 proteins again using the best-fit scaling factor and obtained the R_{hydro} values indicated by the blue circles in Figure 8A; the corresponding experimental values are indicated by green squares. For purposes of

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

comparison with an alternative method of modeling the conformations of the IDPs, we also calculated the average R_{hydro} value of 100 conformations of each IDP generated with the RCG program;[73] these values are indicated by red upward triangles. Looking at the experimental data points, it can be seen that the R_{hydro} values generally increase with the number of residues in the IDP with the notable exception of ProTα (indicated by the asterisk), which despite having only 109 residues has a R_{hydro} larger than CFTR-R, which has 190 residues. Encouragingly, this result is qualitatively reproduced by COFFDROP and is not captured by RCG (Figure 8A). The two sets of computational results are plotted versus the experimental values in Figure 8B; the correlation coefficients with the experimental data points are 0.959 and 0.743 for COFFDROP and RCG, respectively.

DISCUSSION

We have described here the expansion of the COFFDROP force field to include backbone pseudoangle and pseudodihedral potential functions, and a large number of intramolecular nonbonded potential functions necessary for long polypeptide chains to be modeled. We have derived these additional potential functions by applying the IBI method[38,39] to MD data obtained from 441 two-residue peptides reported in recent work[40] and from a further 57 MD simulations of peptide systems reported here. In all cases, we have found that the IBI procedure enables optimized potential functions to be derived that produce a high level of agreement with the probability distributions obtained from MD (see Figures 2, 3, and S3 (Supporting Information)). We have shown that the iterative nature of the IBI method yields potential functions that can differ substantially from those derived by BI of the MD distributions (Figure 4), and we have shown that IBI can refine poor initial guesses to yield robust, optimized potential functions (Figure 5).

Coupling the potential functions derived here for two-residue peptides with those we derived previously for describing interactions of two amino acids,[32] we have a complete, albeit somewhat rudimentary set of potential functions that can be used to simulate entire polypeptide chains. As noted earlier, we make a number of assumptions in applying parameters derived from one- and two-residue peptides to model longer peptides. It is not difficult to imagine, therefore, that there might be problems properly describing longer peptide systems with COFFDROP. Nevertheless, direct tests on a variety of three-residue peptides show that it generally performs very well: only 3 out of the 56 tested peptides produced probability distributions that could be considered to be in poor agreement between MD and BD. We have not been able to identify obvious factors to explain the poor results with these particular systems, but it is perhaps more notable that there are so few failures. It is also worth noting that we have shown that problematic cases can be handled by applying the IBI method directly to three-residue peptides, which means that in principle potential functions need not be derived solely from two-residue peptides, but carrying out a comprehensive parametrization of the force field using MD simulation data on all 9,261 three-residue peptides appears currently infeasible.

In addition to testing COFFDROP's ability to reproduce simulation data, we have tested its ability to reproduce experimental data on the hydrodynamic radii of 8 IDPs. The results that we obtain are, for the most part, quite good. It might at first sight be considered disappointing that the force field could not replicate the experimental R_{hydro} values "out of the box" but instead required a down-scaling of its nonbonded interactions. It should be remembered, however, that COFFDROP has been parametrized to reproduce MD simulation data, not to reproduce experimental data, so any inaccuracies that are present in the MD force field will necessarily also be present in COFFDROP. It is therefore worth noting that Petrov and Zagrovic[78] have recently reported that, in MD simulations of native conformers of the villin headpiece, aggregation appears to occur at concentrations lower than those observed experimentally when the Amber force field used here is employed (aggregation occurs also with a variety of other simulation force fields). We have already shown in our previous work[32] that the nonbonded potential functions in COFFDROP reproduce the erroneous behavior seen in all-atom MD but can be easily adjusted so that they also produce behavior more consistent with the experiment. It is interesting that the nonbonded scaling factor used in that work (0.80) is very similar to the optimal value found here for IDPs (0.825).

An especially encouraging result is that COFFDROP qualitatively reproduces the anomalous R_{hydro} value of ProTα relative to the other IDPs studied here; in contrast, the alternative RCG methodology[73] fails to identify ProTα as unusual (Figure 8A). While COFFDROP produces good qualitative agreement with the experiment, it is important to note that at a quantitative level the agreement is less good. We think that reproducing experimental data for a protein such as ProTα presents a particular challenge to COFFDROP because of its substantial enrichment in like-charged residues: it has a net charge-to-length ratio that is ∼5-fold greater than that of any of the other IDPs simulated. Since COFFDROP's parameters were derived from MD simulations that included no salt, it is likely that the quantitative overestimate of ProTα's R_{hydro} value indicates that the use of a Debye−Hückel model to mimic salt-screening of electrostatic interactions is inappropriate rather than that the force field itself is critically compromised in its ability to model electrostatic interactions. While a fuller exploration of the effects of salts on the conformational behavior of IDPs or other proteins will require that COFFDROP be further extended to enable modeling of a range of salts, the results presented here indicate that in its present form COFFDROP is likely to be useful for modeling the conformational dynamics of unstructured proteins that are not dominated by strong charge−charge interactions.

What we have not done in this study is to develop COFFDROP so that it is capable of simulating and maintaining protein secondary and tertiary structure. We have not seriously pursued the implementation of secondary structures because COFFDROP adopts a deliberately simplified CG mapping scheme in which a single pseudoatom is used to represent the backbone of each residue. This coarseness does not allow COFFDROP to explicitly model intramolecular backbone hydrogen bonding, and it is therefore also unlikely to allow it to directly model the formation and stability of protein secondary structure in a meaningful way.

We note that there are a variety of protein CG force fields in the literature that are able to model protein secondary structure successfully. Importantly, all of these models use finer grain representations of the backbone that allow for intramolecular hydrogen bonding involving the protein backbone. The PACE force field,[79] for example, uses a united atom protein representation that is intended to be used in combination with a CG water model: PACE has been shown to successfully

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

fold a number of small proteins. The PaLaCe force field[80] uses an all-atom representation of the protein backbone and a CG representation of the side chain that involves up to five pseudoatoms. The UNRES[81−83] and FREADY[84] force fields use two pseudoatoms to represent the backbone and a single pseudoatom for the side chain; in these two force fields, secondary structure is maintained by an additional hydrogen bonding potential function that is dependent on the orientation and distance between backbone pseudoatoms. Other more fine-grained CG models for proteins include the PRIMO force field,[31] which uses four pseudoatoms to describe the protein backbone and up to five pseudoatoms for each side chain, the OPEP force field,[30] which uses five pseudoatoms for the backbone and a single pseudoatom for the side chain, and the SIRAH force field,[85] which uses three pseudoatoms for the backbone and up to five for the side chain.

While the MD data that we have used to derive COFFDROP could also be used to derive parameters for a more fine-grained simulation model, a more immediate goal for us is to apply the existing COFFDROP scheme to much larger biomolecular systems.[86,87] With such a goal in mind, the comparatively coarse backbone representation implemented in COFFDROP provides two major advantages: (1) it limits the number of CG beads that are required to model a system, something that can be very important if, for example, hydrodynamic interactions are to be included in the simulations,[88,89] and (2) it allows considerably larger timesteps (50 fs) to be used. For modeling unstructured proteins, we expect the COFFDROP potential functions presented here to be adequate; for modeling more structured proteins, e.g., multidomain proteins that are connected by flexible linkers, it will clearly be necessary to supplement the COFFDROP potential functions with additional terms that stabilize secondary and tertiary structure elements. In our recent simulation study using COFFDROP's nonbonded functions to model intermolecular associations of the villin headpiece,[32] we used Gō-type[90] potential functions to maintain native structure. An obvious alternative, however, would be to use an elastic network model;[91] this has been used successfully in applications of other CG force fields that employ one bead to model the backbone, including the SCORPION[92,93] and MARTINI force fields.[94−96]

■ ASSOCIATED CONTENT

Supporting Information

Images of the CG amino acid models overlaid onto the all-atom models; error in computed R_{hydro} values versus bead radius; comparison of BD and MD probability distributions for nonbonded functions in the Trp-Trp peptide; comparison of potential functions optimized against three independent replicate MD data sets for the Asp-Val peptide; worst predicted results for a three-residue peptide; and computed R_{hydro} values versus nonbonded scaling factor. Tables listing the heavy atoms that are used for the placement of pseudoatoms; experimental details of the simulated IDPs; correlation coefficients between probability distributions sampled from BD and MD of three residue peptides with partial and full IBI optimizations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: adrian-elcock@uiowa.edu.

Funding

This work was supported by NIH R01 GM099865 and R01 GM087290 awarded to A.H.E.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Noid, W. G. Perspective: coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.
(2) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodriguez-Ropero, F.; van der Vegt, N. F. Systematic coarse-graining methods for soft matter simulations - a review. *Soft Matter* **2013**, *9*, 2108−2119.
(3) Saunders, M. G.; Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.
(4) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.
(5) Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137.
(6) Hills, R. D., Jr.; Brooks, C. L., III. Insights from coarse-grained Gō models for protein folding and dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889−905.
(7) Savelyev, A.; Papoian, G. A. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20340−20345.
(8) Korolev, N.; Luo, D.; Lyubartsev, A. P.; Nordenskiöld, L. A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers* **2014**, *6*, 1655−1675.
(9) Maffeo, C.; Ngo, T. T. M.; Ha, T.; Aksimentiev, A. A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment. *J. Chem. Theory Comput.* **2014**, *10*, 2891−2896.
(10) Naômé, A.; Laaksonen, A.; Vercauteren, D. P. A solvent-mediated coarse-grained model of DNA derived with the systematic Newton inversion method. *J. Chem. Theory Comput.* **2014**, *10*, 3541−3549.
(11) Murtola, T.; Falck, E.; Patra, M.; Karttunen, M.; Vattulainen, I. Coarse-grained model for phospholipid/cholesterol bilayer. *J. Chem. Phys.* **2004**, *121*, 9156−9165.
(12) Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.
(13) Lyubartsev, A. P. Multiscale modeling of lipids and lipid bilayers. *Eur. Biophys. J.* **2005**, *35*, 53−61.
(14) Shi, Q.; Izvekov, S.; Voth, G. A. Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane-bound ion channel. *J. Phys. Chem. B* **2006**, *110*, 15045−15048.
(15) Izvekov, S.; Voth, G. A. Solvent-free lipid bilayer model using multiscale coarse-graining. *J. Phys. Chem. B* **2009**, *113*, 4443−4455.
(16) Wang, Z.-J.; Deserno, M. A systematically coarse-grained solvent-free model for quantitative phospholipid bilayer simulations. *J. Phys. Chem. B* **2010**, *114*, 11207−11220.
(17) Hadley, K. R.; McCabe, C. A structurally relevant coarse-grained model for cholesterol. *Biophys. J.* **2010**, *99*, 2896−2905.
(18) Hadley, K. R.; McCabe, C. A coarse-grained model for amorphous and crystalline fatty acids. *J. Chem. Phys.* **2010**, *132*, 134505.
(19) Liu, P.; Izvekov, S.; Voth, G. A. Multiscale coarse-graining of monosaccharides. *J. Phys. Chem. B* **2007**, *111*, 11566−11575.
(20) Srinivas, G.; Cheng, X.; Smith, J. C. A solvent-free coarse grain model for crystalline and amorphous cellulose fibrils. *J. Chem. Theory Comput.* **2011**, *7*, 2539−2548.
(21) Cho, H. M.; Gross, A. D.; Chu, J.-W. Dissecting force interactions in cellulose deconstruction reveals the required solvent versatility for overcoming biomass recalcitrance. *J. Am. Chem. Soc.* **2011**, *133*, 14033−14041.
(22) Markutsya, S.; Kholod, Y. A.; Devarajan, A.; Windus, T. L.; Gordon, M. S.; Lamm, M. H. A coarse-grained model for β-D-glucose based on force matching. *Theor. Chem. Acc.* **2012**, *131*, 1162.

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(23) Markutsya, S.; Devarajan, A.; Baluyut, J. Y.; Windus, T. L.; Gordon, M. S.; Lamm, M. H. Evaluation of coarse-grained mapping schemes for polysaccharide chains in cellulose. *J. Chem. Phys.* **2013**, *138*, 214108.
(24) Srinivas, G.; Cheng, X. L.; Smith, J. C. Coarse-grain model for natural cellulose fribrils in explicit water. *J. Phys. Chem. B* **2014**, *118*, 3026−3034.
(25) Thorpe, I. F.; Zhou, J.; Voth, G. A. Peptide folding using multiscale coarse-grained models. *J. Phys. Chem. B* **2008**, *112*, 13079−13089.
(26) Betancourt, M. R.; Omovie, S. J. Pairwise energies for polypeptide coarse-grained models derived from atomic force fields. *J. Chem. Phys.* **2009**, *130*, 195103.
(27) Wang, Y.; Voth, G. A. Molecular dynamics simulations of polygln aggregation using solvent-free multiscale coarse-grained models. *J. Phys. Chem. B* **2010**, *114*, 8735−8743.
(28) Engin, O.; Villa, A.; Peter, C.; Sayar, M. A challenge for peptide coarse graining: transferability of fragment-based models. *Macromol. Theory Simul.* **2011**, *20*, 451−465.
(29) Terakawa, T.; Takada, S. Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-Terminal domain. *Biophys. J.* **2011**, *101*, 1450−1458.
(30) Sterpone, F.; Nguyen, P. H.; Kalimeri, M.; Derreumaux, P. Importance of the ion-pair interactions in the OPEP coarse-grained force field: parametrization and validation. *J. Chem. Theory Comput.* **2013**, *9*, 4574−4584.
(31) Kar, P.; Gopal, S. M.; Chen, Y.-M.; Predeus, A.; Feig, M. PRIMO: A transferable coarse-grained force field for proteins. *J. Chem. Theory. Comput.* **2013**, *9*, 3769−3788.
(32) Andrews, C. T.; Elcock, A. H. COFFDROP: A coarse-grained nonbonded force field for proteins derived from all-atom explicit solvent molecular dynamics simulations of amino acids. *J. Chem. Theory. Comput.* **2014**, *10*, 5178−5194.
(33) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, Switzerland, 1996; pp 1−1042.
(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.
(35) Hornak, V.; Abel, R.; Okur, A.; Stockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioninf.* **2006**, *65*, 712−725.
(36) 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., Bioninf.* **2010**, *78*, 1950−1958.
(37) 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. Che. Phys.* **2004**, *120*, 9665−9678.
(38) Soper, A. K. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **1996**, *202*, 295−306.
(39) Reith, D.; Pötz, M.; Möller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(40) Li, S.; Andrews, C. T.; Frembgen-Kesner, T.; Miller, M. S.; Siemonsma, S. L.; Collingsworth, T. D.; Rockafellow, I. T.; Nguyet, A. T.; Campbell, B. A.; Brown, R. F.; Guo, C.; Schrodt, M.; Liu, Y.-T.; Elcock, A. H. Molecular dynamics simulations of 441 two-residue peptides in aqueous solution: conformational preferences and neighboring residue effects with the Amber ff99sb-ildn-nmr force field. *J. Chem. Theory Comput.* **2015**, *11*, 1315−1329.
(41) Yang, H.; Elcock, A. H. Association lifetimes of hydrophobic amino acid pairs measured directly from molecular dynamics simulations. *J. Am. Chem. Soc.* **2003**, *125*, 13968−13969.
(42) Thomas, A. S.; Elcock, A. H. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J. Am. Chem. Soc.* **2004**, *126*, 2208−2214.
(43) Thomas, A. S.; Elcock, A. H. Direct observation of salt effects on molecular interactions through explicit-solvent molecular dynamics simulations: Differential effects on electrostatic and hydrophobic interactions and comparisons to Poisson-Boltzmann theory. *J. Am. Chem. Soc.* **2006**, *128*, 7796−7806.
(44) Thomas, A. S.; Elcock, A. H. Direct measurement of the kinetics and thermodynamics of association of hydrophobic molecules from molecular dynamics simulations. *J. Phys. Chem. Lett.* **2011**, *2*, 19−24.
(45) Brown, R. F.; Andrews, C. T.; Elcock, A. H. Stacking free energies of all DNA and RNA nucleoside pairs and dinucleoside-monophosphates computed using recently revised AMBER parameters and compared with experiment. *J. Chem. Theory. Comput.* [Online early access]. DOI: 10.1021/ct501170h. Published Online: Mar, 27, **2015**.
(46) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(47) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(48) Li, D.-W.; Brueschweiler, R. NMR-based protein potentials. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778−6780.
(49) Nosé, S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(50) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(51) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: a new molecular-dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(52) 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*, 8577−8593.
(53) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *12*, 1463−1472.
(54) Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least square procedures. *Anal. Chem.* **1964**, *36*, 1627−1639.
(55) Ermak, D. L.; McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(56) Sangster, M. J. L.; Dixon, M. Interionic potentials in alkali halides and their use in simulations of the molten salts. *Adv. Phys.* **1976**, *25*, 247−342.
(57) Frembgen-Kesner, T.; Elcock, A. H. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory. Comput.* **2009**, *5*, 242−256.
(58) Rotne, J.; Prager, S. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.* **1969**, *50*, 4831−4837.
(59) Yamakawa, H. Transport properties of polymer chains in dilute solution: hydrodynamic interactions. *J. Chem. Phys.* **1970**, *53*, 436−443.
(60) Marsh, J. A.; Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **2010**, *98*, 2883−2390.
(61) Li, Y.; Shan, B.; Raleigh, D. P. The cold denatured state is compact but expands at low temperatures: hydrodynamic properties of the cold denatured state of the c-terminal domain of L9. *J. Mol. Biol.* **2007**, *368*, 256−262.
(62) Guez, V.; Nair, S.; Chaffotte, A.; Bedouelle, H. The anticodon-binding domain of tyrosyl-tRNA synthetase: state of folding and origin of the crystallographic disorder. *Biochemistry* **2000**, *39*, 1739−1747.
(63) Guijarro, J. I.; Pintar, A.; Prochnicka-Chalufour, A.; Guez, V.; Gilquin, B.; Bedouelle, H.; Delepierre, M. Structure and dynamics of the anticodon arm binding domain of *Bacillus stearothermophilus* tyrosyl-tRNA synthetase. *Structure* **2002**, *10*, 311−317.

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(64) Danielsson, J.; Jarvet, J.; Dambert, P.; Graslund, A. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer's Aβ(1−40) peptide. *Magn. Reson. Chem.* **2002**, *40*, S89−97.
(65) Danielsson, J.; Liljedahl, L.; Barany-Wallje, E.; Sonderby, P.; Kristensen, L. H.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E.; Poulsen, F. M.; Maler, L.; Graslund, A.; Kragelund, B. B. The intrinsically disordered RNR inhibitor Sml1 is a dynamic dimer. *Biochemistry* **2008**, *47*, 13428−13437.
(66) Haaning, S.; Radutoiu, S.; Hoffmann, S.; Dittmer, J.; Giehm, L.; Otzen, D. E.; Stougaard, J. An unusual intrinsically disordered protein from the model legume *Lotus japonicus* stabilizes proteins *in vitro*. *J. Biol. Chem.* **2008**, *283*, 31142−31152.
(67) Yi, S.; Boys, B. L.; Brickenden, A.; Konermann, L.; Choy, W.-Y. Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch. *Biochemistry* **2007**, *46*, 13120−13130.
(68) Goldgur, Y.; Rom, S.; Ghirland, R.; Shkolnik, D.; Shadrin, N.; Konrad, Z.; Bar-Zvi, D. Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. *Plant Physiol.* **2007**, *143*, 617−628.
(69) Krishnan, V. V.; Lau, E. Y.; Yamada, J.; Denning, D. P.; Patel, S. S.; Colvin, M. E.; Rexach, M. F. Intramolecular cohesion of coils mediated by phe-gly motifs in the natively unfolded domain of a nucleoporin. *PLoS Comput. Biol.* **2008**, *4*, e1000145.
(70) Uversky, V. N.; Lee, H.-J.; Li, J.; Fink, A. L.; Lee, S.-J. Stabilization of partially folded conformation during α-Synuclein oligomerization in both purified and cytosolic preparations. *J. Biol. Chem.* **2001**, *276*, 43495−43498.
(71) Baker, J. M. R. Structural Characterization and Interactions of the CFTR Regulatory Region. Ph.D. Thesis, University of Toronto, Canada, 2009.
(72) The Uniprot Consortium. Activities at the universal protein resource. *Nucleic Acids Res.* **2014**, *42*, D191−D198.
(73) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099−13104.
(74) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778−795.
(75) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. The determinants of pK_as in proteins. *Biochemistry* **1996**, *35*, 7819−7833.
(76) Li, P. J.; Johnston, H.; Krasny, R. A Cartesian treecode for sceened coulomb interactions. *J. Comput. Phys.* **2009**, *228*, 3858−3868.
(77) Ortega, A.; Amorós, D.; García de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **2011**, *101*, 892−898.
(78) Petrov, D.; Zagrovic, B. Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLoS Comput. Biol.* **2014**, *5*, e1003638.
(79) Han, W.; Schulten, K. Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: improved backbone hydration and interactions between charged side chains. *J. Chem. Theory. Comput.* **2012**, *8*, 4413−4424.
(80) Pasi, M.; Lavery, R.; Ceres, L. PaLaCe: A coarse-grain protein model for studying mechanical properties. *J. Chem. Theory. Comput.* **2013**, *9*, 785−793.
(81) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. Calculation of protein backbone geometry from α-carbon coordinates based on peptide-group dipole alignment. *Protein Sci.* **1993**, *2*, 1697−1714.
(82) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kozlowska, U.; Scheraga, H. A. Parametrization of backbone-electrostatic and multibody contributions to the UNRES force field for protein-structure prediction from ab initio energy surfaces of model systems. *J. Phys. Chem. B* **2004**, *108*, 9421−9438.
(83) Liwo, A.; Khalili, M.; Scheraga, H. A. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362−2367.
(84) Májek, P.; Elber, R. A corase-grained potential for fold recognition and molecular dynamics simulations of proteins. *Proteins: Struct., Funct., Bioninf.* **2009**, *76*, 822−836.
(85) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory. Comput.* **2015**, *11*, 723−739.
(86) McGuffee, S. R.; Elcock, A. H. Diffusion, crowding and protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **2010**, *6*, e1000694.
(87) Trovato, F.; Tozzini, V. Diffusion within the cytoplasm: a mesoscale model of interacting macromolecules. *Biophys. J.* **2014**, *107*, 2579−2591.
(88) Frembgen-Kesner, T.; Elcock, A. H. Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase-barstar association. *Biophys. J.* **2010**, *99*, L75−L77.
(89) Elcock, A. H. Molecule-centered method for accelerating the calculation of hydrodynamic interactions in Brownian dynamics simulations containing many flexible biomolecules. *J. Chem. Theory Comput.* **2013**, *9*, 3224−3239.
(90) Taketomi, H.; Ueda, Y.; Gō, N. Studies of protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **1975**, *2*, 445−459.
(91) Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(92) Ha-Duong, T. Protein backbone dynamics simulations using coarse-grained bonded potentials and simplified hydrogen bonds. *J. Chem. Theory Comput.* **2010**, *6*, 761−773.
(93) Basdevant, N.; Borgis, D.; Ha-Duong, T. Modeling protein-protein recognition in solution using the coarse grained force field SCORPION. *J. Chem. Theory. Comput.* **2013**, *9*, 803−813.
(94) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(95) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved parameters for the Martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.
(96) Stark, A. C.; Andrews, C. T.; Elcock, A. H. Toward optimized potential functions for protein-protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the Martini coarse-grained force field. *J. Chem. Theory Comput.* **2013**, *9*, 4167−4185.

DOI: 10.1021/acs.jctc.5b00038
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX