JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Further Optimization of a Hybrid United-Atom and Coarse-Grained Force Field for Folding Simulations: Improved Backbone Hydration and Interactions between Charged Side Chains

Wei Han and Klaus Schulten*

Beckman Institute and Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: PACE, a hybrid force field that couples united-atom protein models with coarse-grained (CG) solvent (*J. Chem. Theory Comput.* **2010**, *6*, 3373), has been further optimized, aiming to improve its efficiency for folding simulations. Backbone hydration parameters have been reoptimized based on hydration free energies of polyalanyl peptides through atomistic simulations. Also, atomistic partial charges from all-atom force fields were combined with PACE to provide a more realistic description of interactions between charged groups. Using replica exchange molecular dynamics, ab initio folding using the new PACE has been achieved for seven small proteins (16−23 residues) with different structural motifs. Experimental data about folded states, such as their stability at room temperature, melting point, and nuclear magnetic resonance nuclear Overhauser effect constraints, were also well reproduced. Moreover, a systematic comparison of folding kinetics at room temperature has been made with experiments, through standard molecular dynamics simulations, showing that the new PACE may accelerate the actual folding kinetics 5−10-fold, permitting now the study of folding mechanisms. In particular, we used the new PACE to fold a 73-residue protein, α3D, in multiple 10−30 μs simulations, to its native states (C_α root-mean-square deviation of ∼0.34 nm). Our results suggest the potential applicability of the new PACE for the study of folding and dynamics of proteins.

INTRODUCTION

Molecular dynamics (MD) has become an indispensable tool for revealing the molecular mechanism of biological processes.[1−3] Despite ever-growing computing power and better sampling methods,[4−6] all-atom MD simulation cannot be used for the modeling of large biomolecular systems over time scales long enough to be of biological interest. To overcome the challenge, coarse-grained (CG) models, in which multiple atomistic sites are grouped into one site, have been developed for proteins, solvent and membrane, significantly enhancing the speed of simulation.[7−9] However, because of inherent simplification, many CG simulations of proteins rely on information from native protein structures through biased methods such as elastic network models[10] or Gō-type potentials.[11] The CG models are limited, therefore, in their use in studies of folding, aggregation, large conformational changes of proteins, or conformational features of intrinsically disordered proteins.[12]

To alleviate the reliance on native structures, several CG protein models have been parametrized to reproduce peptide folding[13−16] or have been built on knowledge-based statistical potentials,[17−20] achieving even in some cases predictive capabilities.[21−25] Nevertheless, it may be more advantageous to combine atomistic and CG models so that protein models contain enough atomistic detail to obviate the need for information about native structures while environments such as water and membrane are still coarse-grained. Such multiscale approaches have been practiced in numerous studies.[26,27] For example, in an early study,[28] a small portion of a protein was modeled with an all-atom force field while the rest of the protein was coarse-grained with a Gō model. The model developed by Shi and co-workers[29] combined all-atom protein force fields with multiple-site lipid and single-site water. A key element in their model is to parametrize interactions between sites of different resolutions by matching the forces calculated from actual atomistic simulations, allowing the model to maintain the stability of small membrane proteins without any biased constraint. Rzepiela and co-workers[30] have recently proposed a mixed model of all-atom and MARTINI CG force fields,[31,32] the latter of which has proven to be efficient for lipid and solvent simulations.[33,34] In their model, all-atom−CG interactions are replaced by CG interactions through dummy sites, such that the need for reparametrization of cross-resolution terms does not arise. The model has been tested for the case of partitioning of dialanine in different solvents, showing promising results.[30]

In this study, we focus on another hybrid force field, namely PACE,[35] in which a united-atom protein model has been developed in conjunction with MARTINI CG water[35−37] and recently also with some of MARTINI CG lipids.[38] In PACE, cross-resolution parameters were optimized through a thermodynamics-based approach, i.e., through reproducing experimental thermodynamic quantities, which is in line with the parametrization scheme of MARTINI. Moreover, in the context of the hybrid force field, the interactions of atomic sites in proteins were parametrized by using different data as references, such as potential of mean force (PMF) profiles of polar interactions from atomistic simulations and statistical backbone potentials from a Protein Data Bank (PDB) coil

Received: August 7, 2012
Published: October 11, 2012

Table 1. Summary of Peptides and Proteins Studied, Including their Sequences, Melting Temperatures [T_m (Kelvin)], Folded Populations near Room Temperature, and Root-Mean-Square Deviations (rmsd, nanometers) of Representative Structures from Experiments and this Work, as well as Respective Representative Computational Studies by Other Groups

		experiment		current work			
name	sequence	fold %	T_m	fold %	T_m[g]	rmsd	refs
AK17[a]	Ac-$(AAKAA)_3$GY-NMe	30−35[b]		42 ± 5	ND[h,i]		57
Fs	Ac-$A_5(AARAA)_3$-NMe	∼50[b]	∼300	45 ± 2	ND[h,i]		58, 59
GB1p	GEWTYDDATKTFTVTE	∼30[c]	<273	31 ± 9	ND[h,i]	0.11[k]	60−62 21, 63, 64 65, 66
D47P	GEWTYDPATKTFTVTE	∼50[c]	299−311	54 ± 7	300−308	0.12	63
GB1m2	GEWTYNPATGKFTVTE	∼75[c]	318−322	79 ± 7	324−332	0.12	64
Trp cage	NLYIQWLKDGGPSSG RPPPS	∼70[d]	∼315	68 ± 9	∼324	0.13	63, 64, 67 21, 68, 69 70−72 73, 74
BBA5	Ac-YRVpSYDFSRSDEL AKLLRQHAG-NMe	∼20[e]	<273	27 ± 9	ND[h,i]	0.19	75−77
α3D	MGSWAEFKQRLAAIK TRLQALGGSEAELAA FEKEIAAFESELQAY KGKGNPEVEALRKEA AAIRDELQAYRHN		>363[f]		ND[h,j]	0.34	78, 79

[a]Note that as AK17 was used to parametrize the force field, it is not included in the test set. [b]Helical contents measured by CD at 300 K.[48,80] [c]Measured by NMR chemical shifts at 298 K.[53] [d]Estimated from the melting curve at 300 K measured by NMR.[54] [e]Measured by CD and fluorescence experiments at 293 K for its V3Y/F8W mutant.[75] [f]Measured by CD at pH 7.0.[81] [g]Simulated T_m values were determined by surveying all replicas in the REMD simulations for the one in which the probability of folded states is closest to 50%. [h]Not determined. [i]For AK17, Fs, GB1p, and BBA5, T_m values could not be determined because the probability of folded states was lower than 50% for all replicas with temperatures of >300 K. Below 300 K, the CG water tends to freeze,[31] and thus, simulations of protein folding become impossible. [j]It is still difficult to achieve an equilibrium sampling of a protein of such size (73 amino acids), even through REMD, and, therefore, we were not able to determine the T_m for α3D. [k]Calculated on the basis of backbone and C_β atoms for GB1p, D47P, and GB1m2, backbone atoms (3−19) for the Trp cage, backbone atoms (1−22) for BBA5, and C_α atoms for α3D.

library.[39] Comprising such different ways of parametrization, PACE has successfully folded several model peptides, not relying on information about their native structures.[40] However, the stabilities of the native structures were generally underestimated by PACE, suggesting that further improvement of PACE is needed.

Hydration of the backbone is known to be critical for controlling the balance between unfolded and folded states of proteins.[41,42] In PACE, backbone and side chain amide groups shared the same set of hydration parameters to describe interactions between amide groups and CG water, optimized by fitting the hydration free energy of *N*-methylacetamide (NMA).[35] Although NMA was termed a model compound for amide groups in early studies,[43] both theoretical[44−46] and experimental[47] studies have demonstrated that NMA favors hydration too much to furnish a suitable model for an internal amide group in longer polypeptides, because of the well-known end group effects in polymers. It was assumed that in PACE the hydration parameters obtained with NMA could be used for peptide backbone hydration, but such an assumption had not been examined. Therefore, the hydration parameters need to be investigated and, very likely, reoptimized by using model compounds other than NMA. We chose polyalanyl peptides as model compounds; their hydration free energies are experimentally unavailable but can be obtained from atomistic simulations.

We also investigated the parameters for electrostatic interactions. In PACE, electrostatic interactions were modeled by effective potentials[35] other than Coulomb potentials. Such an approach reduced the number of parameters needed for PACE but renders PACE less compatible with MARTINI that uses Coulomb potentials for interactions between charged groups[31] and, thereby, potentially complicates actual parametrization of cross-resolution terms. Also, the physical meaning of the approach is less clear, which may lead to difficulties in interpreting simulation results related to electrostatic interactions. Accordingly, we added Coulomb potentials back to PACE for interactions between charged groups. The parameters for Coulomb potentials, i.e., atomistic partial charges, were directly taken from all-atom force fields, as is done for most of the hybrid force fields.[28−30] The transferability of the atomistic charges was then evaluated through comparison between PMFs from PACE and atomistic simulations.

The new PACE was validated by ab initio folding simulations of a series of peptides and proteins, as summarized in Table 1. These peptides and proteins have a wide spectrum of structural motifs, including α-helix (Fs[48−50]), β-strand (GB1p[51] and its mutants[52,53]), α-PPII (Trp cage[54]), ββα (BBA5[55]), and ααα (α3D[56]) systems, providing a robust test of whether the force field is balanced among different structures. Using an enhanced sampling method, namely replica exchange molecular dynamics (REMD),[4] we were able to perform extensive conformational sampling and, thereby, could compare systematically the simulated results with thermodynamic and structural quantities measured in experiments. We show that the force field has been significantly improved in reproducing experimental observations. Also, standard MD simulations were conducted to

demonstrate the ability of the force field to fold small peptides and proteins at physiological temperature, including the 73-residue α3D. Therefore, our force field is potentially useful for the study of folding and dynamics of whole protein systems.

MODELS AND METHODS

Here we introduce the terms in the PACE force field as well as describe briefly the simulation conditions and the secondary structure analysis applied to folding trajectories.

PACE Force Field. The details of PACE and its parametrization have been discussed in previous studies.[35,37] Only a brief introduction will be given here. PACE preserves all heavy atoms in proteins as well as their backbone and side chain amide hydrogen atoms.[35] The overall potential energy of PACE is expressed as

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} + E_{\phi,\psi,\chi_1} + E_{\text{CGW-CGW}} + E_{\text{CGW-UA}} + E_{\text{vdW}} + E_{\text{polar}} \quad (1)$$

where the first four terms account for generic bonded interactions (E_{local}) mediated through covalent bonds, E_{ϕ,ψ,χ_1} accounts for local interactions involved with rotamers of the backbone (ϕ and ψ) and side chains (χ_1), and the remaining terms account for nonbonded interactions. Specifically, the first four terms describe the energy associated with molecular geometries

$$E_{\text{local}} = \sum_b \frac{1}{2}K_{\text{bond},b}(r_b - r_{0,b})^2 + \sum_a \frac{1}{2}K_{\text{angle},a}(\theta_a - \theta_{0,a})^2 + \sum_d K_{\text{dih},d}[1 + \cos(n_d\zeta_d - \zeta_{0,d})] + \sum_{1-4\text{ pair}} 4\varepsilon_{14,ij}\left(\frac{\sigma_{14,ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{14,ij}^{6}}{r_{ij}^{6}}\right) + \sum_i \frac{1}{2}K_{\text{imp},i}(\xi_i - \xi_{0,i})^2 \quad (2)$$

In eq 2, bonds and angles are governed by harmonic potentials with the following force constants: $K_{\text{bond}} = 1.25 \times 10^5$ kJ nm^{-2}, and $K_{\text{angle}} = 300$ kJ mol^{-1} rad^{-2}. The equilibrium bond lengths (r_0) and angle values (θ_0) were taken from optimized geometries by quantum mechanics (QM) calculations.[37] Dihedral potentials consist of both a cosine function of a dihedral and the Lennard-Jones (LJ) potentials for one to four atom pairs associated with the dihedral. The dihedral parameters were obtained by fitting QM dihedral potential profiles of small molecules.[37] Improper terms, imposing constraints ($K_{\text{imp}} = 300$ kJ mol^{-1} rad^{-2}) on four sites of a dihedral angle, are used to maintain the planarity or chirality of groups.

PACE is designed to provide a correct description of conformational distributions of the backbone (ϕ and ψ) and side chains (χ_1) (Figure 1a) for all amino acids. Previous statistical studies[39,82] have shown that the same amino acid assumes significantly different backbone (ϕ and ψ) conformations for distinctly different side chain conformations, such as *gauche+*, *gauche−*, and *trans* rotamers (Figure 1b), the difference arising from conformationally specific interactions between the side chain and the adjacent backbone. The dependence between backbone and side chain conformations is

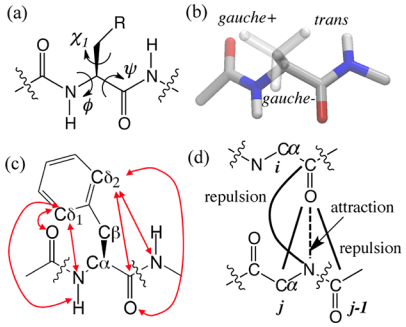

Figure 1. (a) Schematic representation of dihedral angles ϕ, ψ, and χ. (b) Schematic representation of three side chain rotamers: *gauche+*, *gauche−*, and *trans*. (c) Schematic representation of short-range pairs (arrows) in Phe for calculation of E_{ϕ,ψ,χ_1}. (d) Effective interaction potentials for backbone HB, with the dashed line for an attractive potential and the thick lines for the repulsive ones.

accounted for in PACE through the term E_{ϕ,ψ,χ_1},[35] which is expressed as

$$E_{\phi,\psi,\chi_1} = \sum_d \sum_{n_d=1}^{N_d} K_{\text{dih}_{n_d},d}[1 + \cos(n_d\zeta_d - \zeta_{0_{n_d},d})] + \sum_{\text{short}} 4\varepsilon_{\text{short},ij}\left(\frac{\sigma_{\text{short},ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{\text{short},ij}^{6}}{r_{ij}^{6}}\right) \quad (3)$$

As shown in the equation, cosine functions with different multiplicities (n) are applied to dihedral ϕ, ψ, or χ_1 for a more accurate description of dihedral potentials. Furthermore, certain short-range atom pairs, analogous to the pairs with a 1−4 relationship, but separated by more than three covalent bonds, have separate parameters ($\varepsilon_{\text{short},ij}$ and $\sigma_{\text{short},ij}$) for modeling the interactions between a side chain and its adjacent backbone. An example of such pairs in Phe is illustrated in Figure 1c, as indicated by arrows. The parameters $K_{\text{dih},n}$, $\zeta_{0,n}$, $\varepsilon_{\text{short},ij}$, and $\sigma_{\text{short},ij}$ in eq 3 were optimized, through iterative equilibrium simulations,[35] against side chain rotamer distributions and rotamer-dependent backbone conformations from a coil library.[39]

In PACE (eq 1), LJ 12−6 potentials are used for the $E_{\text{CGW-CGW}}$ interactions between CG water, $E_{\text{CGW-UA}}$ interactions between CG water and protein sites, and E_{vdW} interactions between nonpolar protein sites, namely

$$E_{\text{A-B}} = \sum_{i\neq j} 4\varepsilon_{\text{A-B},ij}\left(\frac{\sigma_{\text{A-B},ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{\text{A-B},ij}^{6}}{r_{ij}^{6}}\right) \quad (4)$$

The parameters of MARITINI water for $E_{\text{CGW-CGW}}$ are as follows: $\varepsilon_{\text{CGW-CGW}} = 5.0$ kJ mol^{-1}, and $\sigma_{\text{CGW-CGW}} = 0.47$ nm. The cross-resolution terms $E_{\text{CGW-UA}}$ were optimized by fitting experimental hydration free energies (HFEs) of 35 compounds that cover 11 types of functional groups.[37] E_{vdW} parameters were optimized on the basis of densities of liquid states and free energies of evaporation of eight organic compounds.[35]

PACE handles polar interactions (E_{polar}) between protein sites through a set of effective LJ potentials that are necessary for the directionality of polar interactions.[35] For example, Figure 1d illustrates the interaction scheme of backbone hydrogen bonds (HB), the potential energy functions of which are expressed as

dx.doi.org/10.1021/ct300696c | J. Chem. Theory Comput. 2012, 8, 4413−4424

$$E_{\mathrm{HB,backbone}} = \sum_{|i-j|>2} 4\varepsilon_{\mathrm{attr,O-N}}\left(\frac{\sigma_{\mathrm{O-N}}^{12}}{r_{\mathrm{O-N},ij}^{12}} - \frac{\sigma_{\mathrm{O-N}}^{6}}{r_{\mathrm{O-N},ij}^{6}}\right) + \sum_{|i-j|>2}\left(\frac{C_{\mathrm{rep,O-C\alpha}}}{r_{\mathrm{O-C\alpha},ij}^{12}} + \frac{C_{\mathrm{rep,O-C}}}{r_{\mathrm{O-C},ij}^{12}} + \frac{C_{\mathrm{rep,C-N}}}{r_{\mathrm{C-N},ij}^{12}}\right) \quad (5)$$

where i and j denote residue numbers. The polar interaction parameters were optimized by fitting potential of mean force (PMF) profiles from atomistic simulations.[35] As already mentioned, we remodeled polar interactions between charged groups with Coulomb potentials

$$E_{\mathrm{ele}} = \sum_{i\neq j} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad (6)$$

where q_i and q_j are atomistic charges on atom pair i and j, respectively, r_{ij} is the pair distance, and ε_0 and ε_r are vacuum permittivity and relative permittivity, respectively. ε_r takes the same value (15) that was used for MARTINI.[31]

All the parameter files for the reoptimized PACE force field as well as the original PACE are available at http://www.ks.uiuc.edu/~whan/PACE/.

Simulation Details. NAMD 2.9,[83] with a minor modification for using PACE parameters, was used to conduct folding simulations. Switching functions were applied to Coulomb potentials from 0.0 to 1.2 nm and to LJ potentials from 0.9 to 1.2 nm. We used a 4−5 fs time step through increasing mass of hydrogen atoms to 5 au by repartitioning mass in H-containing groups, which should not affect energy conservation.[84] The temperature was maintained by Langevin dynamics with a damping coefficient of 0.1 ps^{-1}; the pressure was maintained by a Nosé-Hoover Langevin piston barostat[83] with a period of 200 fs and a decay rate of 100 fs. To study the folding of peptides, PDB structures were solvated in cubic boxes of CG waters with 1.5 nm buffer clearance from the solutes, subject to 10 ns heating at 700 K after optimization. The denatured structures, randomly selected from the heating simulations, were used for both standard MD and replica exchange MD (REMD) simulations.[4,85] REMD simulations were performed under constant *NVT* conditions. For each system, 18 replicas, ranging from 300 to 480 K, started with different denatured structures. Attempts to exchange between replicas were made every 2 ps. The exchange ratio between the replica was on average 25−35%. All the folding simulations in this study are summarized in Table 2.

Parametrization of PACE required calculation of hydration free energies (HFE) of solutes and potentials of mean force (PMFs) of interacting pairs. By following previous studies,[35,37] we calculated HFEs and PMFs through thermodynamic integration (TI) methods.[86] GROMACS 3.3[87] was used here to be consistent with the parametrization procedure employed previously for PACE.[35,37] Details about HFE and PMF calculations are presented in Figure S1 and Table S1 of the Supporting Information.

Structure Analysis. Following the work of Sorin and Pande,[59] we considered a residue to be α-helical only if this residue and both of its neighbors have their ϕ and ψ angles within $-60 \pm 30°$ and $-47 \pm 30°$, respectively. Similarly, we considered two residues to be involved in β-sheets if their ϕ and ψ angles are within $-135 \pm 45°$ and $135 \pm 45°$, respectively, and also if there is at least a HB between the backbone amide groups adjacent to the two residues; we note that a HB is considered to be formed when a donor−acceptor distance is shorter than 0.35 nm and a donor−hydrogen−acceptor angle is larger than 120.0°.

An iterative clustering algorithm proposed by Daura et al.[88] was used to group conformations based on the root-mean-square distances of selected atoms. In each iteration, a cluster with the largest number of neighbors that are within a given cutoff of the rmsd of the center of the cluster is removed from a pool of conformations. The remaining pool is used for the next iteration. The algorithm ends when the pool becomes empty.

Table 2. Summary of System Sizes, Simulation Times, and MD Methods Used for Folding Simulations Performed in This Work

system	no. of water molecules	time (μs)[b]	MD method
dipeptide[a]	550	0.05	REMD[c]
AK17	1600	0.2	REMD
Fs	1850	0.2	REMD
	1850	3.0	MD
GB1p	1700	0.75	REMD
D47P	1700	0.35	REMD
	1800	1.3, 2.1, 3.5	MD
GB1m2	1700	0.35	REMD
Trp cage	1800	0.2	REMD
	1800	3, 3	MD
BBA5	1800	0.5	REMD
	1800	8, 8	MD
α3D	4500	10, 20, 30	MD

[a]REMD simulations were conducted for all 20 natural amino acids. [b]For REMD simulations, the simulation time of each replica is listed. For standard MD simulations, the simulation time of each run is listed. Simulation times of multiple runs are all listed, separated by commas. [c]Each REMD involved has 18 replicas with associated temperatures ranging from 300 to 480 K.

RESULTS AND DISCUSSION

In the following sections, we present how we improved the PACE force field by reoptimizing hydration parameters of protein backbones and incorporating Coulomb interactions into the force field. Then we will verify these improvements by examining the performance of the new PACE in reproducing thermodynamics and kinetics in folding processes of a series of small proteins.

Reoptimizing Backbone Hydration Parameters. As models for backbone hydration, blocked polyalanyl peptides (Ace-Ala_n-NMe) have been investigated previously in several computational studies.[44−46] Nevertheless, available HFE data from atomistic simulations were still inadequate for the purpose of parametrization. Therefore, we performed thermodynamic integration (TI) simulations[86] to calculate HFEs of polyalanyl peptides from one to four alanyl groups long, using the OPLS-AA force field[89] and TIP3P water model.[90] An antiparallel β-conformation ($\phi = -139°$; $\psi = 135°$)[46] was chosen as a model for extended and exposed backbones of polypeptides (Table S1 of the Supporting Information).

Figure 2a shows that HFEs of polyalanyl peptides, calculated with the OPLS-AA force field, decrease linearly with the number of alanyl groups. Each individual internal amide group contributes similarly to overall HFEs of polypeptides,

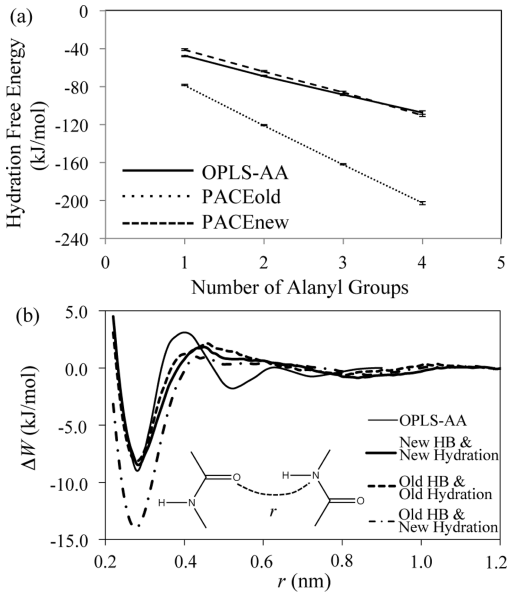

Figure 2. (a) Hydration free energies of extended (ϕ = −139°; ψ = 135°) polyalanyl peptides having one to four alanyl groups, calculated with the OPLS-AA, original PACE (PACEold), and new PACE (PACEnew) force fields. The error bars are estimated with three independent simulations. (b) Potential mean force (PMF) profiles of two interacting amides in water, calculated with the OPLS-AA force field (thin line), the new PACE (thick line), the original PACE (dashed line), and the original HB potentials with the new hydration parameters (dotted−dashed line).

independent of peptide length, consistent with previous computational studies.[45,46,91] The HFE of an internal amide group was estimated, according to an average over all incremental changes of the calculated HFEs, to be −20 ± 2.3 kJ mol^{-1}. Our result is comparable to the results from previous explicit solvent simulations, such as −17 ± 1.3 kJ mol^{-1} for extended polyalanyl peptides[46] and approximately −23 to −25 kJ mol^{-1} for extended polyglycyl peptides.[91,92]

In the original PACE, hydration parameters for amide groups were optimized on the basis of methylacetamide (NMA), producing an HFE of −42 ± 1.5 kJ mol^{-1} for NMA,[35] close to the experimental value (approximately −42 kJ mol^{-1}).[43] However, our calculations (Figure 2a) show that these parameters led to an HFE of −41 ± 1.4 kJ mol^{-1} for internal amide groups in peptide backbones, similar to the HFE of NMA, but significantly stronger than the value of −20 ± 2.3 kJ mol^{-1} for internal amide groups obtained from the atomistic simulations. Several factors may contribute to the discrepancy between HFEs of internal amide groups from the CG and atomistic simulations. (1) Previous Poisson−Boltzmann calculations[45] demonstrated that adjacent amide groups in peptide backbones could affect their electrostatic solvation free energies, significantly reducing the HFE of internal amide groups compared to that of an isolated amide group. However, the CG solvent here is lacking in regard to electrostatic solvation. (2) Detailed amide−water hydrogen bonding (HB), which could also be important for backbone hydration,[46] is missing because of coarse-graining of the water model. (3) Comparative studies between CG and atomistic simulations[93] have shown that coarse-graining could lead to a considerable loss of configurational entropy associated with atomistic details. Thus, HFEs obtained with PACE may also be affected by the loss of entropy because of the CG water model. Considering the factors listed above, it is difficult to adopt only one set of hydration parameters for both internal and isolated amide groups in the current PACE. Therefore, we turned to different parameter sets for the two types of amide groups.

By fitting the HFEs of polyalanyl peptides from the atomistic simulations, we reoptimized the parameters for hydration of internal amide groups, including ε_{CGW-H}, ε_{CGW-O}, σ_{CGW-H}, and σ_{CGW-O}, for the interactions between CG water and backbone nitrogen and hydrogen atoms, assuming $\sigma_{CGW-O} = \sigma_{CGW-H}$ (eq 4) to simplify the optimization. The new parameters are listed in Table S2 of the Supporting Information. The HFEs of polyalanyl peptides, calculated with the new parameters (Figure 2, dashed line), deviate from the all-atom HFEs by ∼3.8 kJ mol^{-1}. With these parameters, the HFE of an internal amide group becomes −23 ± 2.0 kJ mol^{-1}.

We also examined the effect of the change in hydration parameters on HBs between backbone amide groups through comparison of potential of mean force (PMF) profiles of interacting amide groups. As shown in Figure 2b, the interaction strength of backbone HBs, estimated as free energy minimum values in the PMFs, is approximately −8.2 kJ mol^{-1} for the original PACE, close to the value (approximately −9 kJ mol^{-1}) from atomistic simulations. Because the new parameters significantly reduce the magnitude of backbone hydration, the desolvation penalty of backbone HBs should also be reduced. As expected, with the new parameters, the HB strength increases to approximately −14.0 kJ mol^{-1}, indicating that backbone HBs become too strong. Therefore, we need to reoptimize the parameters for backbone HBs by referring to atomistic PMFs. For the same reason, the parameters for HBs involved with backbone amides also need to be examined and, if necessary, reoptimized. We also noticed that the change in the backbone hydration parameters affected backbone conformational distributions. Accordingly, we reoptimized the terms in eq 3 to reproduce statistical ϕ, ψ, and χ_1 distributions from a coil library.[39] The details of all the reoptimizations are discussed in Figures S2 and S3 of the Supporting Information, and the new parameters are summarized in Tables S2 and S3 of the Supporting Information. The new parameters were optimized in the same way as in previous studies,[35] except that parametrization for backbone amide HBs no longer relies on structural features of any specific peptide, but only on an amide−amide PMF obtained through atomistic simulation (Figure S4 of the Supporting Information), thereby rendering the new parameters more general than the original ones.

Incorporation of All-Atom Charges in PACE for Charged Side Chains. We combined atomistic partial charges, taken from OPLS-AA (Figure S5 of the Supporting Information), with PACE to replace effective LJ potentials for interactions between charge pairs. The choice of atomistic force fields is not particularly important as other all-atom force fields, such as CHARMM22[94] and AMBER03,[95] have partial charges for charged groups very close to those of OPLS-AA. To calculate partial charges for PACE, we combined OPLS partial charges on a heavy atom and its attached hydrogen atoms and applied the sum to the corresponding atom in PACE. To see how well PACE and the atomistic partial charges work together, the PMFs of charge pairs, generated by the combined parameters, were compared with respective all-atom PMFs reported by Masunov and Lazaridis.[96]

The comparisons (Figure 3) reveal four key features of charge interactions modeled by the combined parameters. First, the Lys$^+$···Asp$^-$/Glu$^-$ PMF has a contact pair minimum (CPM)

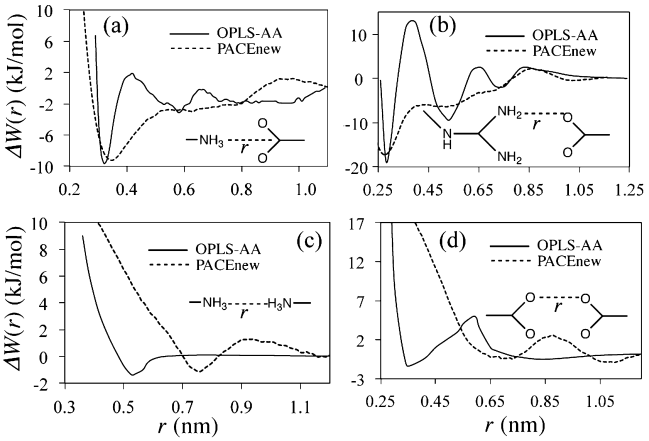

Figure 3. PMFs of $Lys^+\cdots Asp^-/Glu^-$ (a), $Arg^+\cdots Asp^-/Glu^-$ (b), $Lys^+\cdots Lys^+$ (c), and $Asp^-/Glu^-\cdots Asp^-/Glu^-$ (d). The solid lines denote PMFs by the OPLS-AA force field in TIP3P water, and the dashed lines (PACEnew) denote those by PACE with the all-atom charges.

of −9.2 kJ mol^{-1} at r_{N-C} = 0.34 nm, matching the same minimum in the atomistic PMF (−9.6 kJ mol^{-1} at r_{N-C} = 0.32 nm) (Figure 3a). The charges on Lys^+ and Asp^-/Glu^- were repartitioned in several arbitrary ways (Figure S6a of the Supporting Information), but no better match could be obtained with the repartitioned charges than with those from OPLS-AA. Second, the atomistic CPM of $Arg^+\cdots Glu^-/Asp^-$ (Figure S6b of the Supporting Information) could not be reproduced by the combined parameters alone, no matter how the charges on Arg^+ were rearranged, probably because of the lack of details of H atoms on Arg^+. Thus, we turned to a set of effective LJ potentials (Figure S5 and eq S3 of the Supporting Information) that had been used for this charge pair in the original PACE.[35] Together with the atomistic charges, these effective LJ potentials improve the description of the $Arg^+\cdots Glu^-/Asp^-$ CPM (Figure 3b). Third, despite electrostatic repulsion between a like-charge pair, such as $Lys^+\cdots Lys^+$ and $Glu^-/Asp^-\cdots Glu^-/Asp^-$, the combined parameters capture water-mediated minima present in the atomistic PMFs (Figure 3c,d), but the pairs at these minima are 0.2−0.3 nm more distant than those in the atomistic PMFs, because of the large size of CG water. Fourth, at intermediate distances (0.35−0.6 nm), the PACE PMFs, compared to their atomistic counterparts, are more attractive for opposite-charge pairs and more repulsive for like-charge pairs (Figure 3).

As MARTINI CG water is unable to screen electrostatics through dielectric properties, an empirical dielectric constant (ε_r = 15) is often introduced to account for dielectric screening.[31] Also, solvation of charged solutes in CG water is parametrized to partly account for the screening. Considering the difference between the CG and all-atom water models, we do not expect a perfect match between the PMFs from PACE and all-atom simulations. For example, charge-pair interactions are overestimated by PACE when the pairs are separated at intermediate distances. However, even a more sophisticated solvent model, such as the generalized Born (GB^{HCT}) model developed by Hawkins and co-workers,[97,98] renders, for example, $Lys^+\cdots Lys^+$ and $Glu^-/Asp^-\cdots Glu^-/Asp^-$ interactions too repulsive (cf. Figures 13a and 16 of ref 96). Accordingly, we need to further assess the applicability of the combined parameters. In this regard, we note that peptide folding is sensitive to the description of charge−charge interactions; an incorrect description could significantly distort free energy landscapes of peptides, thereby leading to misfolding.[61,62,67,68] Indeed, peptide folding has been recognized as an important opportunity for improvement of force fields such as the GB model.[63,68] As will be shown later, the current charge parameters seem to be applicable for folding simulations, at least for the model peptides studied in this work.

Folding Simulations of Helical Peptides AK17 and Fs.

AK17 is a 17-residue polyalanine-based model peptide (Table 1) known through CD measurements[48] to have 30−35% helical content at 300 K. AK17 has been used to parametrize the backbone HB in the original PACE.[35] A REMD simulation (200 ns/replica) for AK17 shows that the new PACE predicts an overly high helical content (72 ± 2%) at 300 K. To reduce the helical content of peptides, one approach is to adjust ϕ and ψ dihedral potentials to ameliorate the preference for the α region in each residue, as done in previous development of AMBER-GS[58] and AMBER99ϕ[59] for helix folding. However, having been parametrized to fit the coil library, the ϕ and ψ dihedral parameters in the new PACE could not be altered further. We fine-tuned, therefore, the stability of local helical turns by changing the strength of all short-range backbone HBs between residue i and residue i + 3 or i + 4 while leaving the other HB parameters untouched. With a reduction in $\varepsilon_{attr,Oi-Ni+3/i+4}$ (eq 5 and Table S2 of the Supporting Information), the new PACE produces 42 ± 5% helical content for AK17, comparable to experimental data. For the sake of consistency, all folding simulations described below were performed with the same parameters, including the reduced $\varepsilon_{attr,Oi-Ni+3/i+4}$ and all the new parameters discussed in previous sections.

A 20-residue helical peptide, known as Fs peptide[49,50] (Table 1), was used to examine the new PACE through a REMD simulation (200 ns/replica) and a 3 μs standard MD simulation at 300 K. The average helical contents from the 300 K replica and the standard simulation are 42 ± 5 and 45 ± 1.8%, respectively, similar to the value of ∼43% of the original PACE; the result from the new PACE agrees with the ∼50% helical content of Fs found via CD measurements at 300 K.[80] The trajectory (t > 50 ns) from the standard simulation was clustered according to a 0.25 nm cutoff of backbone root-mean-square distance (rmsd) (see Models and Methods). The top five clusters among 547 identified are shown in Figure 4. The

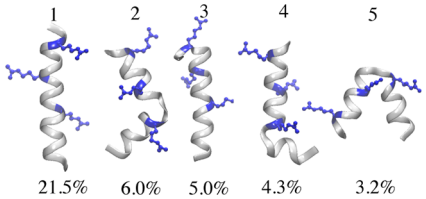

Figure 4. Structures of the first five of the most stable Fs clusters with their respective occupancy indicated below the structures. The structures were drawn with VMD.[113]

most stable cluster (cluster 1, 21.5%) is a single full helix (SFH); in contrast, SFH was only marginally stable (∼1%) in the simulation with the original PACE.[35] On the other hand, the most stable cluster predicted by the original PACE is a reversed helix−turn−helix structure (HTH) with a P_{HTH} of 14.6%,[40] but the same structure is much less stable with the new PACE (cluster 5). The new PACE results agree also more

closely with previous all-atom simulations that showed that Fs has ∼17% SFH.[59]

Because radii of gyration (R_g), which reflect overall shapes of proteins, are experimentally available for both AK17 and Fs,[57,59] we computed protein R_g values for both peptides to examine the ability of the new PACE to describe shapes. The average R_g values for AK17 and Fs are 0.78 ± 0.05 and 0.90 ± 0.06 nm, respectively, close to the experimental values (∼0.82 and ∼0.9 nm, respectively) and the values (0.78 ± 0.06 and 0.85 ± 0.07 nm, respectively) calculated with the original PACE.[35] The analysis of R_g suggests that the shapes of both helical peptides are described reasonably well by both the original PACE and the new PACE.

Folding Simulations of β-Hairpin GB1p and Its Mutants. β-Hairpin GB1p (Table 1), derived from residues 41−56 in the C-terminus of the protein G B1 domain, has been well characterized experimentally.[51−53,99] Because of its fast folding kinetics (τ = 6−20 μs),[53,99] GB1p has also been intensively studied in folding simulations.[21,40,60−64,66] GB1p is known to be marginally stable, having a melting temperature (T_m) of <273 K.[52] The stability of GB1p can be increased to different extents through mutations in its loop such as D47P[53] or the replacement of DDATKT by NPATGK (GB1m2)[52] (Table 1).

We performed REMD simulations for GB1p and its D47P and GB1m2 mutants with the new PACE. We took residues 41−56 of the NMR structure (PDB entry 1gb1) as a reference (Figure 5a) for a folded state. A simulated structure was considered folded if its rmsd of backbone and C_β atoms is less than a certain cutoff. To choose a proper cutoff, we calculated the folded populations of the three peptides by varying the rmsd cutoff from 0.1 to 0.34 nm (Figure S7 of the Supporting Information) and looked for points after which the populations level off. Such points, as shown in the figure, turned out to be at 0.2 nm, which was then chosen as the rmsd cutoff. For GB1m2 and D47P, the folded structures appeared early (t < 50 ns) in the REMD simulations and the population of folded peptides converged after 200 ns (Figure S8 of the Supporting Information). Thus, we extended the simulations of GB1m2 and D47P to 350 ns and extracted the last 150 ns of the simulations for analysis. On the other hand, GB1p was found to fold only after 400 ns; the percentage P_{fold} of folded states converged only after 600 ns. Indeed, a previous folding study with GB solvent involving REMD simulations also found that GB1p folded more slowly than its loop mutants.[63] Garcia and co-workers suggested that there are high barriers dividing folded and unfolded states for GB1p.[100] Therefore, we conducted extended (750 ns) REMD simulations for GB1p and analyzed the last 150 ns of the simulations.

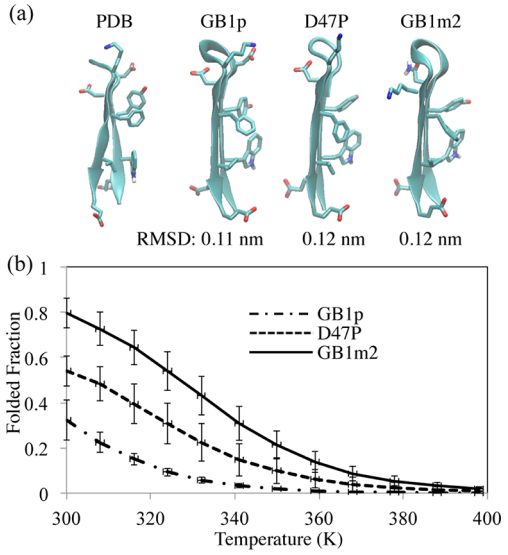

Figure 5. (a) Representative structures of the most stable clusters in simulations of GB1p and its D47P and GB1m2 mutants with their respective RMSD values indicated. (b) Melting curves obtained through REMD simulations of GB1p (dot-dashed), D47P (dashed), and GB1m2 (solid).

At 300 K, the folded populations (P_{fold}) of GB1p, D47P, and GB1m2 were estimated to be 31 ± 9, 54 ± 7, and 79 ± 7%, respectively, in good agreement with P_{fold} values (∼30, ∼50, and ∼75%, respectively) measured at 298 K by NMR experiments.[52,53] In particular, the stability of folded GB1m2 is better reproduced by the new PACE than by the original one (P_{fold} ∼ 40−45%).[40] The clustering (r_{rmsd} = 0.2 nm) results show that all the peptides have their most stable cluster folded (Figure 5a). The other clusters are only marginally stable (<10%). Melting curves were plotted by calculating $P_{fold}(T)$ for all replicas (Figure 5a). The melting curves show that the melting temperature of GB1p should be below 300 K and that the T_m values of D47P and GB1m2 are 300−308 and 324−332 K, respectively, in close agreement with the experimental T_m values (299−311 and 318−322 K, respectively).[52,53] Therefore, the new PACE not only folds GB1p and its mutants into their native structures but also reproduces the delicate stability difference among the peptides. Reproducing relative stabilities of GB1p and its mutants has also been reported by other groups. Chen and Brooks showed that their CHARMM22/GBSW force field was able to differentiate stabilities of GB1p and its β-hairpin-enhancing or -retarding mutants,[63] as did the implicit solvent force field developed by Irback and Mohanty.[64] However, these force fields were optimized purposely for folding of GB1p and its mutants, which is in contrast to the new PACE here that did not incorporate any information about those β-hairpin peptides when it was optimized.

Folding Simulations of Trp cage. Trp cage (Table 1) is a de novo designed 20-residue miniprotein that behaves in several respects like large globular proteins.[54] Trp cage exhibits a compact hydrophobic core, centered around the Trp6 side chain that is caged by side chains of Tyr3, Leu7, Pro12, and Pro18 (Figure 6a). Because of its small size and fast kinetics (τ_f ∼ 4.1 μs),[69] Trp cage has been computationally studied in numerous folding simulations that reproduced its NMR structures with high accuracy.[21,63,64,67−74] The original PACE is also able to fold Trp cage (P_{fold} = 30−35%) with its representative folded structure exhibiting a backbone rmsd of 0.2 nm from the NMR structure,[40] but the original PACE underestimates the P_{fold} of Trp cage compared to the P_{fold} (∼70% with a 0.24 nm rmsd cutoff) measured by CD and NMR experiments.[54]

P_{fold} of Trp cage with the new PACE was calculated through 200 ns REMD simulations, using a 0.24 nm rmsd cutoff of the backbone (residues 3−19). This cutoff was calculated in the same way that was discussed previously (Figure S7 of the Supporting Information). The second halves of simulations, which showed the converged P_{fold} values (Figure S8 of the Supporting Information), were used for analysis. P_{fold} at 300 K was estimated to be 68 ± 9%, consistent with the experimental value (∼70%). We clustered the 300 K ensemble with an rmsd

dx.doi.org/10.1021/ct300696c | J. Chem. Theory Comput. 2012, 8, 4413−4424

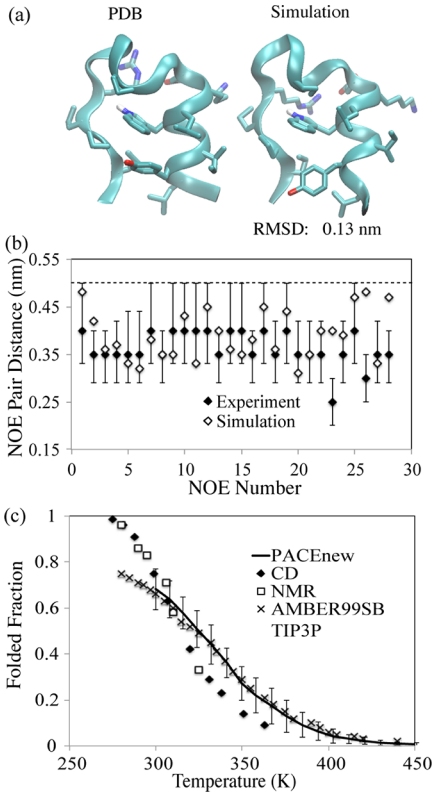

Figure 6. (a) NMR structure of Trp cage (model 1 in PDB entry 1L2Y) and representative structure of the most stable cluster in the 300 K simulation. (b) NOE pair distances from NMR experiments (◆)[54] and the simulation by the new PACE (◇) for long-range pairs ($i/i + n$ with $i \geq 5$). Error bars show reported upper and lower bounds for the NOE distance from NMR measurements. The dashed line denotes a 0.5 nm distance. (c) Folded populations of Trp cage vs temperature, calculated with the new PACE (—) and AMBER99SB with TIP3P water (×) and measured with NMR (◆) and CD (□) methods. Error bars were estimated through block average with 10 ns windows.

cutoff of 0.2 nm. The representative structure of the most stable cluster (69%) has an rmsd of 0.13 nm, which is even better than that (backbone rmsd ~ 0.16 nm) from previous OPLS-AA/SPC simulations.[74]

The melting curve of Trp cage is plotted in Figure 6b. For comparison, we also plotted the melting data from NMR and CD experiments,[54] and from previous ~40 μs REMD simulations with AMBER99SB[101] and TIP3P water for converged sampling.[70] As shown in Figure 6c, the melting curve generated by the new PACE agrees closely, for the entire temperature range of the REMD simulations, with that from the all-atom simulations. The T_m measured from the PACE simulations is ~324 K, close to ~321 K from the all-atom simulations, and slightly higher than the experimental value (315 K), but much lower than those from other all-atom simulations with OPLS-AA in TIP3P (440 K)[74] and AMBER94[102] in GB^{HCT} (~400 K).[67] Both the AMBER99SB melting curve and ours are in overall agreement with the experimental melting data but depart noticeably (~15% folded population) from the experimental data for extreme temperatures.

We also calculated interproton NOE pair distances and compared them with the reported data from NMR experiments.[54] We first added all missing H atoms back to each frame of the simulations and then evaluated NOE pair distances as $R_{NOE} = \langle R_{HH}^{-6} \rangle^{-1/6}$, which was averaged over the whole ensemble (t > 100 ns). As suggested by Zhou,[74] a simulated NOE pair was considered to disagree with the NOE observations if its distance is ≥0.025 nm below the lower bound or beyond the upper bound in the NOE distance range.[54] For 83% of 169 reported NOE pairs, the simulated distances are within the NOE distance range, comparable to the results from OPLS-AA/SPC simulations in which 92% of NOE pairs showed correct distances.[74] Of the 17% of the pairs in violation, only three had a distance of >0.5 nm, a typical distance limit for observing NOE signals. We further looked into 28 long-range NOE pairs that are separated by at least five residues ($i/i + n$ with $n \geq 5$). These long-range NOE pairs provided key constraints for defining topology and tertiary structures of Trp cage.[54] As shown in Figure 6b, only three of 28 long-range pairs violate NMR NOE distances, while there were also four long-range pair violations in the OPLS-AA/SPC simulations.[74] The three pairs with violations include (1) the pair between Y3 HD1 and W6 HH2, suggesting weaker packing between Y3 and W6 (Figure 6a), (2) pairs between W6 HZ2 and P12 HA, and (3) pairs between D9 HB1 and S14 HB2, which have also violated NOE distances in the OPLS-AA/SPC simulations.[74] Despite the three distance violations, all the long-range pair distances are still within the 0.5 nm limit. Therefore, we conclude that the PACE simulations reproduced well the NMR structures for Trp cage.

Folding Simulations of BBA5. De novo designed BBA5 (Table 1) is characteristic of globular proteins, namely having a well-defined tertiary structure composed of all types of elementary structures, including a β-hairpin (residues 1−8), an α-helix (residues 12−20), and a connecting loop (Figure 7a).[55,103] Folding of BBA5 has been achieved in all-atom simulations with both explicit[76] and implicit[75,77,104] solvent models. We examined folding of BBA5 through REMD simulations with the new PACE. To achieve a converged sampling, the simulation were carried out for 500 ns (Figure S8 of the Supporting Information), the last 300 ns of which were taken for analysis.

We tried to use backbone rmsd to define folded states for calculation of P_{fold}, but we were unable to choose a definite rmsd cutoff as P_{fold} was found to increase linearly with the rmsd cutoff, without any inflection point (Figure S7 of the Supporting Information), unlike in the case of GB1p or Trp cage. Thus, we set the cutoff to 0.3 nm, a value used also in previous simulations of BBA5,[76,77] and P_{fold} was calculated to be 27 ± 9% at 300 K. A BBA5 double mutant with enhanced stability, as shown in previous CD and fluorescence experiments,[75] has a P_{fold} of only ~20%. Because the folded structures are marginally stable in the simulation, it is necessary to know if the lowest free energy minimum is still folded. The simulated ensemble was therefore clustered using a 0.25 nm backbone rmsd cutoff. The most stable cluster (28%) turned out to be folded, as reflected by a low rmsd (0.19 nm) for its representative structure (Figure 7a).

The β-hairpin and α-helix in BBA5, on the other hand, show considerable stabilities. The probability P_β of the β-hairpin is 77%. The helical content P_α of residues 12−18 is 80%. The high P_β and P_α values agree with values determined in previous folding simulations of BBA5 with OPLS/GBSA.[104] Our results reveal that a large fraction of unfolded states have considerable native secondary structures. However, the hairpin adopts various orientations relative to the helix in the unfolded states,

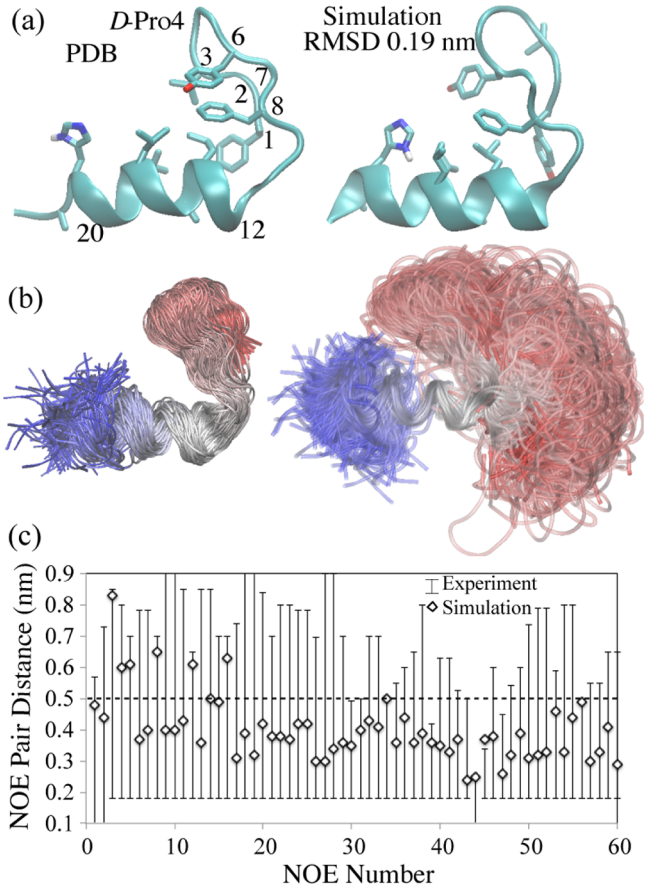

Figure 7. (a) NMR structure of BBA5 (PDB entry 1T8J) and representative structure of the most stable cluster in the 300 K simulation. (b) Overlay of randomly selected conformations from the most stable cluster (left), superimposed with backbone residues 1−21, or from the entire simulations (right), superimposed with backbone residues 12−19. (c) NOE pair distance upper and lower bounds from NMR experiments (vertical line)[54] and simulated NOE distances by the new PACE (◇) for long-range pairs ($i/i + n$ with $i \geq 5$). The dashed line denotes 0.5 nm.

as demonstrated by a structural overlay of randomly selected conformations from the entire ensemble (Figure 7b). Previous simulations had shown that even at 278 K the relative orientation of the hairpin in the folded states is flexible.[75] NMR studies have also demonstrated that certain mutations on the BBA5 hairpin could alter the relative orientation of the hairpin or even lead to multiple structures with different hairpin orientations.[55]

To further validate the marginal stability of BBA5 and its flexibility, we calculated NOE pair distances for the entire ensemble and compared them with NMR data.[55] Of 347 reported NOE pairs, five pairs violate NMR NOE constraints. Of 60 long-range ($i/i + n$ with $n \geq 5$) NOE pairs, only one pair (R2 HA and D7 HB*) violates its constraint, showing a calculated NOE distance 0.03 nm above the experimental upper bound (0.34 nm) for this pair (Figure 7c). We noticed that the good agreement between the simulated and experimental NOE distances may be partly due to large upper bounds of certain experimental NOE constraints, some of which are even beyond 0.8 nm. To see if the good agreement relies on those large upper bounds, we also compared all the NOE pair distances with the 0.5 nm distance that is the typical distance limit for observing NMR signals. The comparison showed that 25 of 347 NOE pair distances exceed the 0.5 nm distance limit. Among the 25 NOE pairs, only six involve long-range NOE pairs, mainly between Y1 and L14 and between Y1 and A15. The majority of the experimental constraints are, according to the calculation, predicted to be observable in the NOE spectrum. Therefore, the conformational ensemble of BBA5 generated by the new PACE reproduces the NMR structural data.

Folding Simulations of α3D. α3D (Table 1) is a de novo designed three-helix-bundle protein that contains 73 residues.[56,105] Because of its length, folding of α3D requires tremendous conformational sampling such that ab initio folding with atomistic detail represents a great computational challenge. Indeed, computational folding of α3D has been achieved in only a few studies. Using an AMBER96/GBSA model, Ozkan and co-workers folded α3D with a C_α rmsd of 0.46 nm by combining a conformational search algorithm with REMD.[78] Recently, Lindorff-Larsen and co-workers, using a modified CHARMM force field,[106] achieved ab initio folding of α3D multiple times in a simulation extending beyond 700 μs.[79] As a final test, we employed the new PACE to fold α3D at 310 K. Starting with denatured α3D, we observed folding of α3D in three extended (10−30 μs) standard MD simulations. The average C_α rmsd of representative folded structures was 0.34 ± 0.02 nm for the three simulations, close to the result (0.31 nm) achieved by Lindorff-Larsen and co-workers.[79]

Our α3D folding simulations lead to several interesting results. First, unfolded states of α3D were found to be disordered with the radius of gyration exhibiting large fluctuations (Figure 8c and Figure S9 of the Supporting Information), yet these states exhibit significant (~25%) residual helical content. A residual helical content of 20% had also been reported for unfolded α3D in the atomistic simulations.[79] Second, for the three helices in α3D, namely,

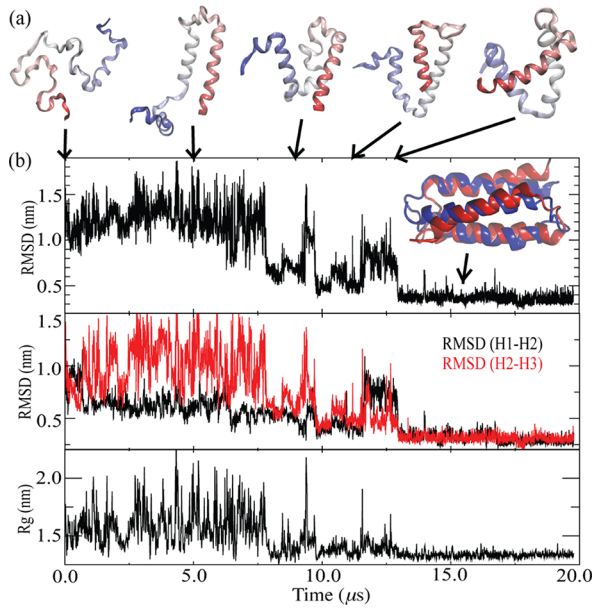

Figure 8. (a) Snapshots of the folding simulation of α3D. (b) C_α rmsd from the native structure (PDB entry 2A3D) (top), C_α rmsds of the H1−H2 complex (black) and H2−H3 complex (red) (middle), and radii of gyration (bottom) of α3D in the folding simulations. Five fraying residues in each terminus were discarded when the C_α rmsd was calculated. A superposition of the representative folded structure (red) and the native structures (blue) is shown in the top panel.

dx.doi.org/10.1021/ct300696c | J. Chem. Theory Comput. 2012, 8, 4413−4424

H1 (residues 1−18), H2 (residues 28−44), and H3 (residues 54−73), H1 and H2 form the native two-helix bundle earlier than H2 and H3 do, as reflected by our rmsd calculation of H1−H2 and H2−H3 complexes (Figure 8b and Figure S9 of the Supporting Information). Our results are in contrast to those of a previous folding study[107] that, using an Ising-like Gō model, suggested that H3 involves the native topology first. Third, in two of the simulations, folding was complete after H3 packed itself to the preformed H1−H2 complex with correct native contacts (Figure S9 of the Supporting Information); in a third simulation emerged intermediates with nativelike topologies but a non-native hydrophobic core [8 μs < t < 13 μs (Figure 8b)], which eventually transited to the folded structure. Such a scenario of multiple folding pathways is expected for α3D according to experiments.[105]

Folding Kinetics with the New PACE. Because several peptides have been folded successfully with the new PACE, it is natural to inspect the kinetics of the simulated folding. For this purpose, additional multimicrosecond simulations were conducted for D47P, Trp cage, and BBA5. Alongside the simulations for Fs and α3D, these standard MD simulations, summarized in Table 2, cover all the structural motifs investigated in this work. The folding time (τ_f) was estimated as the average time span of folding starting when the simulations began or when the peptides unfolded and ending when the peptides reached folded states (Figures S9−S13 of the Supporting Information). As shown in Table 3, the simulated τ_f values are 5−10-fold shorter than the experimental ones, except for the case of α3D, which is probably due to different pH conditions, i.e., neutral pH in our simulations and pH 2.2 in the experiments.[105] An acidic pH can lower the T_m of α3D from >363 to 343 K;[105] recent single-molecule FRET experiments also show that the folding kinetics of α3D around room temperature becomes ∼10-fold faster at low pH compared to neutral pH.[108] Compared to the τ_f from the atomistic simulations (27 ± 8 μs),[79] our τ_f is 2−3 times shorter (11 ± 4 μs), though it was measured at a lower temperature (310 K vs 370 K). Altogether, for the systems tested here, folding modeled by the new PACE is ∼5−10-fold faster than that seen in experimental measurements.

Table 3. Number of Folding Events (N_f) in the Standard MD Simulations and Comparison of Folding Time $\tau_{f,sim}$ (microseconds) with Experimental $\tau_{f,expt}$

peptide[a]	N_f	$\tau_{f,sim}$	$\tau_{f,expt}$	$\tau_{f,expt}/\tau_{f,sim}$
Fs	45	0.04 ± 0.01	0.16−0.22[b]	∼4.5−6.5
D47P	4	0.9 ± 0.3	∼13[c]	∼15
Trp cage	7	0.4 ± 0.13	4.1[d]	∼11
BBA5	13	0.8 ± 0.3	8 ± 3.5[e]	∼9
α3D	3	11 ± 4	3 ± 1.2[f]	∼0.3

[a]Standard MD simulations were performed at 300 K for Fs, D47P, and Trp cage and at 310 K for BBA5 and α3D. [b]At 301 K with T-jump CD.[109,110] [c]At 298 K with NMR.[53] [d]At 296 K with T-jump CD.[69] [e]At 298 K with T-jump fluorescence.[75] [f]At 323 K and pH 2.2 with T-jump IR.[105]

The 5−10-fold acceleration of folding by the new PACE could be attributed to multiple factors, such as coarse-graining of the solvent model and simplified potentials for protein−protein and protein−solvent interactions, both smoothing energy landscapes of folding. For instance, the CG water model is less viscous and allows solutes to diffuse 4−5 times faster than atomistic models, as demonstrated in previous studies.[31] However, the observed factor of acceleration of folding may not be enough for estimation of the time scale of CG simulations because protein dynamics includes events on multiple time scales, each of which could be affected to a different extent through coarse-graining of atomistic models. Also, we have redistributed masses of hydrogen-containing groups for a larger time step, which could affect hydrogen bonding dynamics[84] and, thereby, complicate the meaning of the time scale of simulations. Therefore, one should be cautious in using the observed acceleration factor for folding in interpretations of the time scale of CG simulations.

CONCLUSION

We have presented a further optimization of the PACE multiscale model and validated it through folding several proteins. We demonstrated that the original PACE overestimates backbone hydration and render small proteins, such as GB1m2, a GB1p mutant, and Trp cage, less stable than they are in experiments. The new PACE, optimized for HFEs of the peptide backbone, reproduces correct stabilities not only for GB1m2 and Trp cage but also for other proteins, such as GB1p and its D47P mutant and BBA5, bringing calculated T_m values and NOE distances into agreement with experiment. We also explored the possibility of combining atomistic charges from atomistic force fields with PACE for interactions between charged groups. We showed that with only minor modification of PACE, the combined parameters recall satisfactory PMFs for charge pairs and work well in folding simulations. We expect that the treatment of electrostatic interactions will be significantly improved by further combining atomistic charges, the new PACE, and recently developed polarizable CG water models.[111,112] The new PACE accelerates folding of the tested proteins 5−10-fold, as demonstrated in a systematic comparison between simulations and experiments. Finally, we were able to fold the 73-residue α3D into its native structure through standard MD simulations, indicating the potential of the new PACE for the study of protein folding and dynamics of large proteins.

ASSOCIATED CONTENT

Supporting Information

Examples of HFE calculations (Figure S1), results of optimized backbone potentials (Figures S2 and S3), amide−amide PMFs through atomistic and PACE simulations (Figure S4), scheme of charge parameters (Figure S5), PMFs of opposite-charge pairs with different parameters (Figure S6), probability of folded structures with different rmsd cutoffs (Figure S7), P_{fold} against t in REMD simulations (Figure S8), folding and unfolding events in standard MD simulations of α3D, D47P, BBA5, Trp cage, and Fs (Figures S9−S13, respectively), summary of HFE calculations (Table S1), new parameters for polar interactions involving backbone and backbone dihedral potentials (Tables S2 and S3), and a discussion on HFE and PMF calculations, optimization of backbone potentials, and LJ parameters of charge-pair interactions. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: kschulte@ks.uiuc.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by National Institutes of Health Grants 9P41GM104601, P41-RR005969, and R01-GM067887 and National Science Foundation Grant MCB02-34938. Computer time was provided by Texas Advanced Computing Center through Grant MCA93S028 allocated by the Extreme Science and Engineering Discovery Environment program funded by the National Science Foundation.

■ REFERENCES

(1) Chen, J.; Brooks, C. L., III; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140−148.
(2) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glaettli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064−4092.
(3) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(5) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9−12.
(6) Bussi, G.; Gervasio, F. L.; Liao, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(7) *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Eds.; CRC Press: Boca Raton, FL, 2008; pp 1−4.
(8) Clementi, C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 10−15.
(9) Klein, M. L.; Shinoda, W. *Science* **2008**, *321*, 798−800.
(10) Tama, F.; Brooks, C. L., III. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 115−133.
(11) Taketomi, H.; Ueda, Y.; Go, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−459.
(12) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(13) Bureau, T.; Deserno, M. *J. Chem. Phys.* **2009**, *130*, 235106.
(14) Das, P.; Matysiak, S.; Clementi, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10141−10146.
(15) Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394−408.
(16) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616−11629.
(17) Korkut, A.; Hendrickson, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15667−15672.
(18) Mukherjee, A.; Bhimalapuram, P.; Bagchi, B. *J. Chem. Phys.* **2005**, *123*, 14901.
(19) Majek, P.; Elber, R. *Proteins* **2009**, *76*, 822−836.
(20) Buchete, N. V.; Straub, J. E.; Thirumalai, D. *Curr. Opin. Struct. Biol.* **2004**, *14*, 225−232.
(21) Chebaro, Y.; Dong, X.; Laghaei, R.; Derreumaux, P.; Mousseau, N. *J. Phys. Chem. B* **2009**, *113*, 267−274.
(22) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849−873.
(23) Fujitsuka, Y.; Chikenji, G.; Takada, S. *Proteins* **2006**, *62*, 381−398.
(24) Ding, F.; Tsao, D.; Nie, H. F.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010−1018.
(25) Hills, R. D., Jr.; Lu, L.; Voth, G. A. *PLoS Comput. Biol.* **2010**, *6*, e1000827.
(26) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.
(27) Nielsen, S.; Bulo, R.; Moore, P.; Ensing, B. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12401−12414.
(28) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys. Rev. Lett.* **2005**, *95*, 218102.
(29) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045−15048.
(30) Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10437−10448.
(31) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(32) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(33) Kasson, M. P.; Kelley, N. W.; Singhal, N.; Vrljic, M.; Brunger, A. T.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11916−11921.
(34) de Vries, A. H.; Yefimov, S.; Mark, A. E.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5392−5396.
(35) Han, W.; Wan, C.-K.; Jiang, F.; Wu, Y.-D. *J. Chem. Theory Comput.* **2010**, *6*, 3373−3389.
(36) Han, W.; Wu, Y.-D. *J. Chem. Theory Comput.* **2007**, *3*, 2146−2161.
(37) Han, W.; Wan, C.-K.; Wu, Y.-D. *J. Chem. Theory Comput.* **2008**, *4*, 1891−1901.
(38) Wan, C.-K.; Han, W.; Wu, Y.-D. *J. Chem. Theory Comput.* **2012**, *8*, 300−313.
(39) Jiang, F.; Han, W.; Wu, Y.-D. *J. Phys. Chem. B* **2010**, *114*, 5840−5850.
(40) Han, W.; Wan, C.-K.; Wu., Y.-D. *J. Chem. Theory Comput.* **2010**, *6*, 3390−3402.
(41) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623−16633.
(42) Bolen, D. W.; Rose, G. D. *Annu. Rev. Biochem.* **2008**, *77*, 339−362.
(43) Wolfenden, R. *Biochemistry* **1978**, *17*, 201−204.
(44) Avbelj, F.; Luo, P.; Baldwin, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10786−10791.
(45) Avbelj, F.; Baldwin, R. *Proteins* **2006**, *63*, 283−289.
(46) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins* **2004**, *55*, 502−507.
(47) Auton, M.; Bolen, D. W. *Biochemistry* **2004**, *43*, 1329−1342.
(48) Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413−8421.
(49) Lockhart, D. J.; Kim, P. S. *Science* **1992**, *257*, 947−951.
(50) Lockhart, D. J.; Kim, P. S. *Science* **1993**, *260*, 198−202.
(51) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584−590.
(52) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238−7243.
(53) Olsen, K. A.; Fesinmeyer, R. M.; Stewart, J. M.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15483−15487.
(54) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(55) Struthers, M. D.; Ottesen, J. J.; Imperiali, B. *Folding Des.* **1998**, *3*, 95−103.
(56) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5486−5491.
(57) Zagrovic, B.; Jayachandran, G.; Millett, I. S.; Doniach, S.; Pande, V. S. *J. Mol. Biol.* **2005**, *353*, 232−241.
(58) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *99*, 2782−2787.
(59) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472−2493.
(60) Zagrovic, B.; Sorin, E. J.; Pande, V. S. *J. Mol. Biol.* **2001**, *313*, 151−169.
(61) Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777−12782.
(62) Zhou, R. H. *Proteins* **2003**, *53*, 148−161.
(63) Chen, J.; Im, W.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2006**, *128*, 3728−3736.
(64) Irback, A.; Mohanty, S. *Biophys. J.* **2005**, *88*, 1560−1569.
(65) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins* **2007**, *66*, 846−859.
(66) Shao, Q.; Gao, Y. Q. *J. Chem. Theory Comput.* **2010**, *6*, 3750−3760.
(67) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587−7592.
(68) Geney, R.; Layten, M.; Gomperts, R.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 115−127.

(69) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(70) Day, R.; Paschek, D.; Garcia, A. E. *Proteins* **2010**, *78*, 1889−1899.
(71) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548−14549.
(72) Simmerling, C. L.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.
(73) Schug, A.; Herges, T.; Wenzel, W. *Phys. Rev. Lett.* **2003**, *91*, 158102/1−158102/4.
(74) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.
(75) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102−106.
(76) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6456−6461.
(77) Jang, S.; Kim, E.; Pak, Y. *J. Chem. Phys.* **2008**, *128*, 105102.
(78) Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11987−11992.
(79) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(80) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200−9210.
(81) Walsh, S. T.; Sukharev, V. I.; Betz, S. F.; Vekshin, N. L.; DeGrado, W. F. *J. Mol. Biol.* **2001**, *305*, 361−373.
(82) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1−102.
(83) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(84) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786−798.
(85) Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. *Chem. Phys. Lett.* **2001**, *335*, 435−439.
(86) van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 171−176.
(87) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(88) Daura, X.; van Gunsteren, W.; Mark, A. E. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 269−280.
(89) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(90) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(91) Hu, C. Y.; Kokubo, H.; Lynch, G. C.; Bolen, D. W.; Pettitt, B. M. *Protein Sci.* **2010**, *19*, 1011−1022.
(92) Gu, W.; Rahi, S.; Helms, V. *J. Phys. Chem. B* **2004**, *108*, 5806−5814.
(93) Baron, R.; de Vries, A. H.; Hunenberger, P. H.; van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 8464−8473.
(94) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evansack, 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., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(95) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(96) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722−1730.
(97) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(98) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(99) Munõz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(100) Garcia, A. E.; Sanbomatsu, K. Y. *Proteins* **2001**, *42*, 345−354.
(101) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(102) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(103) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **1996**, *118*, 3073−3081.
(104) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775−786.
(105) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C.-Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486−15491.
(106) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47−L49.
(107) Zamparo, M.; Pelizzola, A. *J. Chem. Phys.* **2009**, *131*, 035101.
(108) Chung, H. S.; Gopich, I. V.; McHale, K.; Cellmer, T.; Louis, J. M.; Eaton, W. A. *J. Phys. Chem. A* **2011**, *115*, 3642−3656.
(109) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691−697.
(110) Thompson, P. A.; Munõz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J. *J. Phys. Chem. B* **2000**, *104*, 378−389.
(111) Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, e1000810.
(112) Wu, Z.; Cui, Q.; Yethiraj, A. *J. Phys. Chem. B* **2010**, *114*, 10524−10529.
(113) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.