New-Generation Amber United-Atom Force Field

Lijiang Yang,[†] Chun-hu Tan,[†] Meng-Juei Hsieh,[†] Junmei Wang,[‡] Yong Duan,[§] Piotr Cieplak,[||] James Caldwell,[⊥] Peter A. Kollman,[#] and Ray Luo*,[†]

Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697, Encysive Pharmaceuticals Inc., Houston Texas 77030, Genome Center, University of California, Davis, California 95616-8816, Burnham Institute for Medical Research, La Jolla, California 92037, Department of Chemistry, Stanford University, Stanford, California 94305, and Department of Pharmaceutical Chemistry, University of California, San Francisco

Received: January 9, 2006; In Final Form: May 11, 2006

We have developed a new-generation Amber united-atom force field for simulations involving highly demanding conformational sampling such as protein folding and protein−protein binding. In the new united-atom force field, all hydrogens on aliphatic carbons in all amino acids are united with carbons except those on Cα. Our choice of explicit representation of all protein backbone atoms aims at minimizing perturbation to protein backbone conformational distributions and to simplify development of backbone torsion terms. Tests with dipeptides and solvated proteins show that our goal is achieved quite successfully. The new united-atom force field uses the same new RESP charging scheme based on B3LYP/cc-pVTZ//HF/6-31g** quantum mechanical calculations in the PCM continuum solvent as that in the Duan et al. force field. van der Waals parameters are empirically refitted starting from published values with respect to experimental solvation free energies of amino acid side-chain analogues. The suitability of mixing new point charges and van der Waals parameters with existing Amber covalent terms is tested on alanine dipeptide and is found to be reasonable. Parameters for all new torsion terms are refitted based on the new point charges and the van der Waals parameters. Molecular dynamics simulations of three small globular proteins in the explicit TIP3P solvent are performed to test the overall stability and accuracy of the new united-atom force field. Good agreements between the united-atom force field and the Duan et al. all-atom force field for both backbone and side-chain conformations are observed. In addition, the per-step efficiency of the new united-atom force field is demonstrated for simulations in the implicit generalized Born solvent. A speedup around two is observed over the Duan et al. all-atom force field for the three tested small proteins. Finally, the efficiency gain of the new united-atom force field in conformational sampling is further demonstrated with a well-known toy protein folding system, an 18 residue polyalanine in distance-dependent dielectric. The new united-atom force field is at least a factor of 200 more efficient than the Duan et al. all-atom force field for ab initio folding of the tested peptide.

1. Introduction

Molecular simulations are now widely applied for investigating structures and functions of biomolecules. In a typical simulation run, a biomolecule is represented by a classical all-atom model, where all intramolecular and intermolecular interactions including those with solvent molecules are described by a molecular mechanics force field.[1−5] However, for simulations involving highly demanding conformational sampling, such as protein folding and protein−protein binding, reduced models are common because energy calculations are more efficient and reduction in degrees of freedom renders conformational sampling much less demanding. These reduced models range from very fast models with one point per residue (either on lattice[6] or off lattice[7]) to other more complex models[8] that provide a more detailed representation for protein structures and dynamics. An added advantage of these reduced models is the absence of explicit representation of solvent molecules, whose contributions have been implicitly wrapped into their model parameters.[9]

Existing reduced models for proteins are mostly developed from nonredundant protein structure databases, thus they are termed knowledge-based potentials.[9] This is in contrast to physics-based potentials, that is, all-atom molecular mechanics force fields, which are developed with respect to quantum mechanical calculations and experimental properties of small molecules.[9] Arguably, physics-based potentials are better than knowledge-based potentials in capturing detailed interactions in proteins, though at the cost of being less efficient. In this work, we explore an alternative route to develop reduced protein models by following the same physical principle with which typical all-atom molecular mechanics force fields are developed. Thus, the reduced protein models are physics-based potentials but they are more efficient than typical physics-based potentials—all-atom molecular mechanics force fields. To achieve the goal for a reduced protein model to offer as much physics as an all-atom molecular mechanics force field, we propose to parametrize the reduced protein model so that its potential energy surface (with respect to the reduced degrees of freedom) is as

* To whom correspondence should be addressed. E-mail: rluo@uci.edu.
[†] UC-Irvine.
[‡] Encysive Pharmaceuticals.
[§] UC-Davis.
[||] Burnham Institute for Medical Research.
[⊥] Stanford University.
[#] Deceased, formerly at UC-San Francisco.

10.1021/jp060163v CCC: $33.50
Published on Web 06/15/2006

close to that of the all-atom force field as possible. That is to say, the reduced protein model is built to be coupled to the all-atom force field from which it is derived for enhancing conformational sampling. There are several advantages in adopting this strategy to develop a reduced protein model: (1) it takes less time to parametrize the model as long as the all-atom model has been parametrized; (2) it also takes less time to make subsequent refinement of the model as long as the all-atom model has been refined first, for example, in the refinement of the notoriously difficult backbone torsion terms; and (3) since a consistency is enforced between the reduced protein model and the all-atom model, it is more straightforward to investigate the efficiency gain in conformational sampling through comparative simulations in the two protein models. As a first step in this direction, we have updated the united-atom model for proteins in the Amber force fields. As will be shown below for a toy protein folding system, even with reduction as modest as the united-atom model, the efficiency gain in conformational sampling is already exciting.

The idea of using united-atom models for efficient simulations goes back to the 1970s when Dunfield et al. developed the UNICEPP force field.[10] In UNICEPP, nonpolar hydrogen atoms are not represented explicitly but are included implicitly by representing nonpolar carbons and their bonded hydrogens as a single particle.[10] Compared with all-atom models, the advantages in using united-atom models are apparent even if only raw efficiency gain in simulations is considered (i.e., how many CPU hours are needed to simulate how long a trajectory). First, they can significantly reduce the size of most problems, since roughly half of the atoms in biological or other organic macromolecules are hydrogens. Thus, there are fewer nonbonded interactions and internal degrees of freedom in united-atom models. Second, larger dynamics integration step sizes can be used by not including hydrogens since their small mass requires a smaller time step for accurate integration. In addition, the positions of hydrogens do not have to be generated, which are usually not available from experimental methods such as X-ray crystallography. However, an often overlooked and more important advantage in adopting united-atom models is the efficiency gain in conformational sampling. It should be pointed out that these advantages become less apparent when the systems are solvated in explicit solvent. Therefore, our primary motivation to develop the new united-atom force field is for applications in implicit solvents due to these solvent models' increasing success and popularity, even if they are still in active development.[11−16]

Earlier comparisons between the all-atom and the united-atom simulations show that the united-atom force field is a satisfactory representation of internal vibrations and bulk properties of small molecules and short peptides.[10,17] However, limitations were also revealed in previous studies:[17] (1) explicit representation of hydrogens was found to be necessary for accurate treatment of hydrogen bonding; (2) π-stacking could not be represented without including hydrogens in aromatic groups explicitly; (3) dipole and quadrupole moments were found inaccurate when uniting hydrogens with polar heavy atoms. New approaches were found to overcome the limitations of united-atom force fields. For example, only aliphatic hydrogens, which are not significantly charged and do not participate in hydrogen bonds, are represented as united atoms while other hydrogens are represented explicitly. In this way, the limitations of the united-atom force field are partially mitigated while preserving most of the benefits of the united-atom force field. Of course, the larger dynamics time step can no longer be used due to the use of polar and aromatic hydrogens. However, with increasing computing power, a factor of about 2 saving from using a larger time step becomes less important.

In contrast, all atoms including all hydrogens are represented explicitly in all-atom force fields. Most previous efforts have been made on constructing both the energy function forms and parameters of all-atom force fields since the early days of biomolecular simulations.[18] Several fully functional force fields for biomolecules were introduced as early as 1980. For example, the Weiner et al. force field is one of the most widely used first-generation all-atom force fields.[1] The Weiner et al. force field used electrostatic potential (ESP) calculated by a quantum mechanical method to derive charges. A hybrid OPLS/Weiner et al. force field that combines OPLS nonbonded parameters[19] and Weiner et al. covalent-bonded parameters was also created and used in many protein systems.[20,21]

A decade later, based on the OPLS philosophy of balanced solvent−solvent and solute−solvent interactions[19] and the Weiner et al. strategy of obtaining parameters through high-level quantum mechanical calculations on dipeptide fragments,[1,22] Cornell et al. developed one of the second-generation Amber force fields[2] at roughly the same time when newly improved OPLS, MMFF, and CHARMM force fields were released.[3−5] Since there are considerable variations among charges generated using different conformations of a molecule in standard ESP, Cornell et al. applied a two-stage RESP charge-fitting method.[23] In addition to the new point charges, a new simplistic van der Waals scheme with parameters mostly borrowed from OPLS was developed to reproduce liquid properties.[2] The Cornell et al. force field has been applied to a wide range of simulations including both nucleic acids and proteins.

Recently, Duan et al. introduced a new all-atom force field for simulations of proteins.[24] Unlike previous all-atom force fields that were based on gas-phase quantum mechanical calculations, atomic charges in the Duan et al. force field were based on B3LYP/cc-pVTZ//HF/6-31g** quantum mechanical electrostatic potentials calculated with the PCM continuum solvent[25,26] in a low dielectric to mimic an organic solvent dielectric environment similar to that of the protein interior.[24] The use of a continuum solvent in quantum mechanical calculations provides an opportunity to polarize molecular wave functions to a desired amount with the aim of balancing protein−protein and protein−solvent interactions because the solvent dielectric can be adjusted freely in the continuum solvent.[24] In this way, one of the limitations of the Cornell et al. force field—lacking a controlled representation of the polarization effect in the condensed phase—can be partially resolved, though still not in a perfect fashion.

Closely following our effort in the Duan et al. force field, the new-generation Amber united-atom force field is developed based on the same high-level quantum mechanical calculations in the PCM continuum solvent.[25,26] In our force field design, only aliphatic hydrogens are united while other hydrogens—polar and aromatic hydrogens—are still represented explicitly. Further, aliphatic hydrogens on the Cα atoms are also represented explicitly to minimize the perturbations to the protein backbone interactions, that is, the same atomic charges and backbone torsion terms can be used in the united-atom force field. On the basis of this design, atomic partial charges are derived with the RESP approach.[2,27] van der Waals parameters are reoptimized starting from published values.[1,17,19,22,28] Parameters for all new torsion terms on the side chains are also refitted based on the new partial charges and van der Waals

parameters. Side-chain torsion and van der Waals parameters other than the ones involving aliphatic carbons and other parameters, including bond and bond angle, are all retained from the Duan et al. force field parameter set.[27] The suitability of existing force field terms with the new nonbonded parameters is also discussed. In the following, all fitting methods are described in section 2. Fitting data and testing results are reported in section 3. Concluding remarks are presented in section 4.

2. Method

2.1. Description of the Model. The new united-atom force field is based on the same effective two-body additive formalism as its all-atom counterparts in Amber. Its bond and angle terms are modeled by harmonic potentials, torsion terms are represented with Fourier series, van der Waals terms are based on 6-12 potentials, and electrostatic terms are computed with Coulomb's law

$$E_{\text{total}} = \sum_{\text{bonds}} K_b(b - b_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right]$$

Here, K_b, b, and b_{eq} are, respectively, the force constant, bond length, and equilibrium value for the bond stretching terms; K_θ, θ, and θ_{eq} are, respectively, the force constant, bond angle, and equilibrium value for the angle bending terms; n, V_n, ϕ, and γ are, respectively, the periodicity, force constant, torsion angle, and phase angle for the torsion terms. The last terms in the above equation are for nonbonded interactions, including van der Waals terms and electrostatic terms. A_{ij} and B_{ij} are van der Waals parameters for the atom pair i and j; q_i and q_j are, respectively, charges for atoms i and j. ϵ is the dielectric constant that takes into account the effect of the medium that is not represented and is usually set to 1.0 in a typical solvated system when solvent is represented explicitly or implicitly.

In this formalism, van der Waals and electrostatic interactions are only calculated between atoms in different molecules and between atoms in the same molecule but separated by at least three bonds. Nonbonded interactions separated by exactly three bonds ("1−4 interactions") are scaled down by a scaling factor. For consistency, the 1−4 scaling factors adopted here are identical to those in the existing all-atom additive force fields in Amber. Thus, the 1−4 van der Waals terms are divided by 2.0 and the 1−4 electrostatic terms are divided by a factor of 1.20.

As outlined in the Introduction, all hydrogen atoms on aliphatic carbon atoms in all amino acids are united with carbon atoms except those on Cα in the new united-atom force field. Our choice of explicit representation of all protein backbone atoms aims at minimizing the perturbation of the united-atom approximation to the backbone interactions and to simplify the development of backbone torsion terms. In doing so, identical partial charges on the backbone as the all-atom force field can be enforced in the two-stage RESP charging scheme.[23] Further, we plan to use the same backbone torsion terms as its all-atom counterpart as will be discussed below. This further reduces the perturbation to the backbone terms.

2.2. Atom Types. Although the atom type of aliphatic carbon atoms in CH, CH_2, and CH_3 groups is CT in the Amber all-atom force fields, these carbon atoms should be regarded as three different types in a united-atom force field because they have different numbers of hydrogen atoms. Thus, three new atom types, C1, C2, and C3, are introduced to represent the united carbon atoms in CH, CH_2, and CH_3, respectively (see Table 1). Their atomic masses are also adjusted according to the number of their bonded hydrogen atoms. The aliphatic carbon atoms and their new types in all natural amino acids are shown in Table S-1 of the Supporting Information.

TABLE 1: Atom Types and New van der Waals Parameters of United-atom Carbons and All-atom Aromatic Carbons, Nitrogens, and Hydrogens, and Hydroxyl Oxygens

type	description	mass (au)	r_m, Å	ϵ, kcal/mol
C1	sp^3 carbon with one hydrogen	13.018	1.9580	0.0994
C2	sp^3 carbon with two hydrogens	14.026	2.0580	0.1094
C3	sp^3 carbon with three hydrogens	15.034	2.0580	0.1494
CU[a]	aromatic carbon atom types	12.010	1.9080	0.1032
NU[b]	aromatic nitrogen atom types	14.010	1.8240	0.2040
H	H bonded to NU	1.008	0.6000	0.0188
H4	H bonded to CU with 1 electrwd. group	1.008	1.4090	0.0180
H5	H bonded to CU with 2 electrwd. groups	1.008	1.3590	0.0180
OH	oxygen in hydroxyl group	16.000	1.6349	0.2104

[a] CU stands for aromatic carbons: CA, CB, CC, CN, CR, CV, CW, and C*. [b] NU stands for aromatic nitrogens: NA and NB.

2.3. RESP Charges. In the united-atom force field, effective atom-centered point charges were obtained by fitting to the quantum mechanically derived electrostatic potentials. In the quantum mechanical calculations, density functional theory (B3LYP) with a basis set of cc-pVTZ was used.[24] This method was shown to reproduce gas-phase dipole moments within 5% for a range of organic molecules.[29] Another important feature of the quantum mechanical calculation is that it is calculated with a continuum solvent to mimic the low dielectric protein interior. This feature allows that the polarization effect induced by the protein environment be considered reasonably in quantum mechanical calculations based on model small organic molecules. In these calculations, each amino acid was represented by a dipeptide fragment consisting of the amino acid residue and two Amber terminal groups (ACE and NME). The electrostatic potential of each dipeptide was calculated for two conformations with backbone torsion angles constrained to $(\varphi,\psi) = (-60,-40)$ and $(\varphi,\psi) = (-120,140)$, respectively, representing the α and β conformations.[2,24]

A two-stage RESP fitting was used in the charge-fitting process as in other Amber force fields.[23] In the first stage, the charge of each atom was free to change under the constraint that the net charge equals to 0, +1(LYS, ARG, HIP), or −1(ASP, GLU, CYM). Two conformers of each dipeptide were considered simultaneously. In the second stage, charges of aliphatic hydrogen atoms were set to zero and the chemically equivalent atoms were set to have the same values, while the charges of the terminal groups and those of backbone peptide bonds were fixed. Just as in Duan et al., we did not enforce the same backbone charges among different dipeptides,[24] so that each amino acid has its own backbone charges. For neutral amino acids excluding PRO, the partial charges (in atomic units) for Cα are in the range of −0.273 to 0.173; those of Hα are in the range of −0.072 to 0.169; those of C are in the range of 0.521 to 0.670; those of O are in the range of −0.596 to −0.543; those of N are in the range of −0.521 to −0.245; and those of H are in the range of 0.243 to 0.305.

2.4. van der Waals Parameters. To investigate the compatibility of the new charging scheme with the existing all-atom van der Waals parameters in Amber, we compared the qualities

of absolute solvation free energies by both Cornell et al. and Duan et al. all-atom force fields for 13 neutral amino acids side-chain analogues (excluding ALA, GLY, and PRO): ASN, CYS, GLN, HIE, ILE, LEU, MET, PHE, SER, THR, TRP, TYR, and VAL. In this study, the absolute solvation free energies were computed only for the Cornell et al. force field, as described below. The absolute solvation free energies for the Duan et al. force field were derived from those of the Cornell et al. force field and the relative solvation free energies between Duan et al. and Cornell et al. The absolute solvation free energies of the new united-atom force field were also investigated and were computed in the same manner as those of the Duan et al. force field. In this comparative study, we found it necessary to empirically adjust the van der Waals parameters for aromatic groups and hydroxyl groups to achieve a comparable agreement with experiment between the Duan et al. and Cornell et al. force fields. The quality of the new parameters is discussed in the Results section.

In the absolute solvation free energy simulations, side-chain analogues were solvated by TIP3P water[30] with a buffer thickness of 11.0 Å. Before free energy simulations were started, solvated systems were fully relaxed with the PMEMD program in Amber8[31] until no systematic drift was observed in running-averaged potential energies (usually less than 4 ns). The absolute solvation free energies were computed by the "annihilation" process:[32] first by decoupling the Coulombic interaction and then by decoupling the van der Waals interaction between solute and solvent. In the first stage, λ was changed from 0 to 1 in the step of 0.04. For each λ, 40 ps molecular dynamics of equilibration was simulated, followed by another 40 ps of data collection. To guarantee convergence in free energies, the simulation times (both equilibration and production runs) were doubled until the free energies differ by less than 0.20 kcal/mol. It was found that 80 ps for equilibration and 80 ps for production were sufficient to satisfy the convergence criterion for decoupling the Coulombic interaction after these solvated neutral side-chain analogues were extensively preequilibrated. In the second stage, the van der Waals interaction between solute and solvent was decoupled by increasing λ from 0 to 0.5 in the step of 0.025 followed by increasing λ from 0.5 to 1 in the step of 0.0125. The uneven step sizes for λ were used to improve the quality in numerical integrations for free energies. It was found that 160 ps for equilibration and 160 ps for production were sufficient to satisfy the convergence criterion (0.20 kcal/mol) for decoupling the van der Waals interaction of the tested neutral side-chain analogues. The above two decoupling processes were also performed in a vacuum to compute the absolute solvation free energies of the tested side-chain analogues. All above simulations were performed in the constant temperature (300 K) and constant pressure (1 bar) ensemble using the Berendsen's coupling algorithms[33] in Amber8.[31] The time step was chosen as 1 fs in the leapfrog[34] numerical integrator for all free energy simulations. Particle mesh Ewald (PME)[35] with default parameters in Amber8[31] was used to treat long-range electrostatics, except that the real-space cutoff for electrostatics and van der Waals interactions was set to be 9 Å. A continuum correction for van der Waals interactions outside the cutoff distance was also used in PME in Amber8.[35]

The van der Waals parameters for the new atom types C1, C2, and C3, introduced in the united-atom force field were refitted starting from the values from previous liquid-state Monte Carlo simulations.[1,17,19,22,28] Specifically, we took into account the van der Waals parameters for the united-atom aliphatic carbon in OPLS.[19] Furthermore, empirical adjustments of r_m and ϵ were employed to reduce the difference in the solvation free energies between the new united-atom force field and the Duan et al. all-atom force field for 18 amino acids (excluding ALA, GLY, and PRO) side-chain analogues: ARG, ASN, ASP, CYS, GLN, GLU, HIE, HIP, ILE, LEU, LYS, MET, PHE, SER, THR, TRP, TYR, and VAL. These solvated systems were prepared and preequilibrated in the same manner mentioned above for absolute solvation free energy simulations. Here, the Duan et al. all-atom force field was used to produce the initial state. Then, from this initial state, the charges of each side-chain analogue were perturbed to our new united-atom charges while λ was changed from 0 to 1 in the step of 0.04. A time of 40 ps was used for equilibration and another 40 ps was used for data collection. The relative free energies calculated at the first stage are pure electrostatic contribution. Second, aliphatic hydrogen atoms were made to disappear and aliphatic carbon atoms change to the new united carbon atoms. At the second stage, 60 λ values (from 0 to 0.5 in the step of 0.025 and from 0.5 to 1 in the step of 0.0125) were used to perturb into the final state, the united-atom charges, and van der Waals parameters. Here, 40 ps was used for equilibration and 40 ps was used for data collection at each λ. The relative free energies calculated at the second stage include both electrostatic and nonelectrostatic contributions. To guarantee convergence in the computed relative free energies, the simulation times (both equilibration and production runs) were doubled until the relative free energies differ by less than 0.20 kcal/mol. It was found that 80 ps for equilibration and 80 ps for production were already sufficient to satisfy an even more stringent criterion (0.01 kcal/mol). After the two stages of perturbation, all the amino acids side-chain analogues with the Duan et al. all-atom force field are perturbed to their final states: those with the united-atom force field charges and new united-atom van der Waals parameters. As reference, relative free energies by perturbing the Cornell et al. all-atom force field to the united-atom force field or to the Duan et al. all-atom force field were also computed.

The standard errors were used as statistical uncertainties for all free energy simulations.[36] For each window, the standard error of $\langle(\mathrm{d}H/\mathrm{d}\lambda)\rangle$ is $\delta\langle(\mathrm{d}H(\lambda)/\mathrm{d}\lambda)\rangle_\lambda = (\sigma/(N_e)^{1/2})$. N_e is the "independent" sampling number, which can be calculated by the correlation time, τ, of $\langle(\mathrm{d}H/\mathrm{d}\lambda)\rangle$ and the data collection time T_{dc}: $N_e = T_{dc}/(2\tau)$.[36] For the side-chain analogues tested here, $\tau < 0.3$ ps.

2.5. Side-Chain Torsion Parameters. Due to the use of new point charges and new van der Waals parameters, side-chain torsion parameters need to be reoptimized. Parameters for each side-chain torsion angle were optimized through a fit between the conformational energy profiles of the united-atom force field and the Duan et al. all-atom force field. Here, a 12-point (every 30 °) potential energy profile was used. Torsion parameters were adjusted to minimize the root-mean-squared deviation of conformational energies between the united-atom calculation and the all-atom calculation.

2.6. Molecular Dynamics Simulations. To examine the overall accuracy of the new united-atom approximation and the force field parameters, molecular dynamics (MD) simulations were performed on three globular proteins: all-α homeodomain (1enh), all-β B domain of protein G (1pgb), and α/β SH3 domain (1shg). Explicit TIP3P solvents were added to fully solvate the proteins in truncated octahedral periodic boxes in the Leap module of Amber8.[31] The minimum distance from the protein surface to the box boundary was set to be 10 Å. Seven Cl^-, one Cl^-, and four Na^+ ions, respectively, were added to

1enh, 1shg, and 1pgb to neutralize the proteins. PME[35] with default parameters in Amber8[31] was used to treat long-range electrostatics, except that the real-space cutoff for electrostatics and van der Waals interactions was set to be 9 Å. MD simulations were started after a brief steepest descent minimization of 1000 steps to relax any possible clashes. SHAKE[37] was turned on for bonds containing hydrogen atoms,[11,15] so that a time step of 2 fs could be used in the leapfrog[34] numerical integrator for MD simulations. Constant temperature (300 K) and constant pressure (1 bar) were maintained by Berendsen's methods.[33] To study the quality of the new united-atom force field vs that of the Duan et al. all-atom force field, a cumulative 80 ns was run (10 independent trajectories of 8 ns each) for each protein in each force field. In total, 480 ns trajectories were collected in the comparative analysis of two force fields with three proteins.

To examine the per-step efficiency of the new united-atom force field in implicit solvents, Langevin dynamics (LD) simulations in the generalized Born (GB) implicit solvent were also performed for the three proteins. The igb option of 5 in Amber 8 was used in the test runs.[11,15] The cavity radii for GB are optimized in a previous study.[38] The cutoff distances were set to the default values of 25 Å for both the self and interaction terms in GB. For timing analysis, 1000 steps of LD were performed for both the new united-atom and Duan et al. all-atom force fields.[24] The timing analysis was performed on a quiet Dell PowerEdge server with dual Intel Xeon CPUs (2.8GHz/1MB L2 Cache/800 MHz Front Side Bus) and 3GB main memory (DDR 400 MHz). A single CPU was used for all LD runs.

2.7. Ab Initio Folding Simulations. To examine the additional efficiency gain in conformational sampling through reduced degrees of freedom in the new united-atom force field, we performed two sets of ab initio folding simulations in the new united-atom force field and Duan et al. all-atom force field, respectively. In this preliminary analysis, we chose a well-defined nontrivial toy system, an 18 residue polyalanine peptide (ALA18, blocked with ACE and NME) in a distance-dependent dielectric ($\epsilon = r$), which is known to fold into a stable helical structure. As reference, two sets of equilibrium simulations (in both the united-atom and the all-atom force fields) starting from the folded helical structure were also performed. In each of the four sets of simulations, 100 ns and at least 10 independent trajectories were performed for a reasonable assessment of the efficiency gain in conformational sampling.

LD simulations at 300 K with a low friction constant ($\gamma = 1$ ps^{-1}) were started after a brief steepest descent minimization of 1000 steps to relax any possible clashes. SHAKE[37] was turned on for bonds containing hydrogen atoms, so that a time step of 2 fs could be used in the leapfrog[34] numerical integrator for all LD simulations.

3. Results

3.1. RESP Charges and van der Waals Parameters. The new united-atom point charges of all natural amino acids in the Amber force field database are shown in Table S-2 in the Supporting Information. The quality of the new point charges due to the united-atom approximation was assessed by comparing dipole moments with those derived from the Duan et al. all-atom force field using the same charging method. Figure 1 shows that dipole moments of all dipeptides computed with the new charges are almost the same as those computed with the Duan et al. all-atom charges. The slope of the trend line is 0.998. This indicates that long-range electrostatic properties of our united-atom charges are very similar to those of the all-atom charges given that the same charging scheme is used.

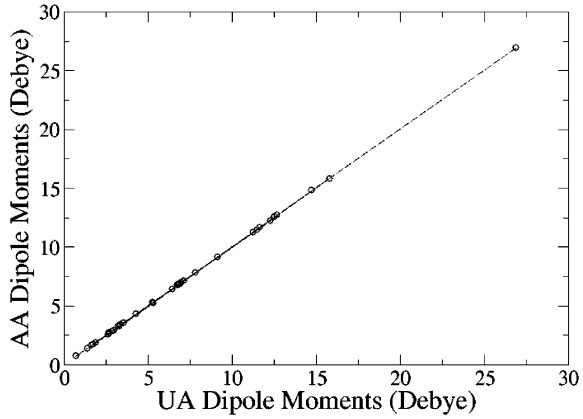

Figure 1. Correlation of dipole moments of all dipeptides in charge fitting between the Duan et al. all-atom (AA) force field and the Yang et al. united-atom (UA) force field in this work.

TABLE 2: Computed Solvation Free Energies with Uncertainties (kcal/mol) of 13 Neutral Amino Acid Side-chain Analogues (excluding ALA, PRO, and GLY) in Duan et al. and Cornell et al. All-atom Force Fields and the Yang et al. United-atom Force Field (this work) Are Compared with Experimental Values from Wolfenden et al. (ref 39)

side-chain analogue	Duan et al.	Yang et al.	Cornell et al.	experiment
ASN	−8.90 ± 0.26	−8.94 ± 0.28	−9.26 ± 0.25	−9.68
CYS	0.23 ± 0.21	0.32 ± 0.23	0.13 ± 0.21	−1.24
GLN	−8.81 ± 0.28	−8.91 ± 0.31	−10.07 ± 0.27	−9.38
HIE	−8.37 ± 0.30	−8.46 ± 0.31	−9.01 ± 0.28	−10.27
ILE	2.37 ± 0.30	2.57 ± 0.35	2.37 ± 0.29	2.15
LEU	2.36 ± 0.30	2.33 ± 0.35	2.29 ± 0.29	2.28
MET	0.39 ± 0.29	0.83 ± 0.32	0.73 ± 0.28	−1.48
PHE	−0.37 ± 0.31	−0.56 ± 0.33	−0.12 ± 0.30	−0.76
SER	−4.52 ± 0.21	−4.16 ± 0.22	−4.46 ± 0.20	−5.06
THR	−4.12 ± 0.26	−4.28 ± 0.28	−3.85 ± 0.24	−4.88
TRP	−5.64 ± 0.34	−5.75 ± 0.36	−4.96 ± 0.33	−5.88
TYR	−4.38 ± 0.34	−4.42 ± 0.35	−4.48 ± 0.32	−6.11
VAL	2.20 ± 0.28	2.32 ± 0.31	2.30 ± 0.26	1.99
rmsd[a]	1.05	1.12	1.06	N.A.

[a] rms deviations from experimental values.

Empirical adjustment of the van der Waals parameters for aromatic and hydroxyl groups is necessary to achieve a comparable agreement with experiment[39] between the Duan et al. charging scheme and the Cornell et al. charging scheme. These new van der Waals parameters, shared between the new united-atom force field and Duan et al. all-atom force field, are listed in Table 1. The quality of the new van der Waals parameters for the Duan et al. charging scheme is tested by comparing simulated absolute solvation free energies with experiment for 13 neutral amino acid side-chain analogues.[39] These data, along with those for the Cornell et al. charging scheme, are shown in Table 2. It can be seen that the Duan et al. force field performs comparably with the Cornell et al. force field after revision of the van der Waals parameters as far as solvation free energies of side-chain analogues are concerned, that is, the root-mean-squared (rms) error is 1.05 kcal/mol for the Duan et al. force field vs 1.06 kcal/mol for the Cornell et al. force field. The new united-atom force field also performs similarly with the two tested all-atom force fields using the new van der Waals parameters, with an rms error of 1.12 kcal/mol.

The need for readjustment of van der Waals parameters, mostly in ϵ values for aromatic groups and hydroxyl groups, indicates a systematic difference between the Duan et al.

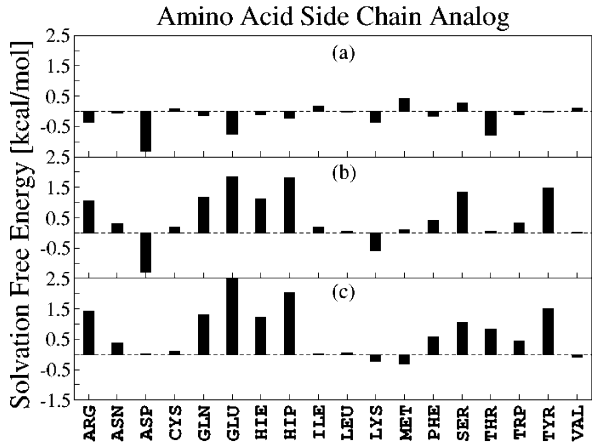

Figure 2. Relative solvation free energies of 18 amino acids side-chain analogues (excluding ALA, GLY, and PRO). Maximum uncertainty is 0.04 kcal/mol for all relative solvation free energies: (a) from the Duan et al. all-atom force field to Yang et al. united-atom force field (this work); (b) from the Cornell et al. all-atom force field to Yang et al. united-atom force field (this work); (c) from the Cornell et al. all-atom force field to Duan et al. all-atom force field.

charging scheme and Cornell et al. charging scheme. This is even so given that the overall similarity in dipole moments between the two charging schemes is very high—with a slope of the fitting trend line of 1.01 reported previously.[24] With such a high similarity in dipole moments between the two charging schemes, it was assumed that the van der Waals parameters could be transferred from the Cornell et al. force field to the Duan et al. force field without change.[24] However, the Duan et al. charging scheme with the Cornell et al. van der Waals parameters gives an rms error of 1.58 kcal/mol in the tested neutral amino acid side-chain analogues. This is 0.52 kcal/mol higher than that of the Cornell et al. charging scheme. Thus, it is difficult to determine which charging scheme is better with the current simplistic van der Waals scheme in Amber. The reason is simple: whichever charging scheme that is in a closer agreement to OPLS would agree better with experiment because OPLS van der Waals parameters and point charges were developed to be consistent with each other. It is subject to debate that there is an optimal partition between van der Waals dispersion and Coulobmic interactions without clear guidance from high-level quantum mechanical calculations in the condensed phase. The two terms apparently compensate with each other in a force field. Indeed, due to this concern, it has been proposed within the Amber community that the simplistic van der Waals parameters for each charging scheme have to be optimized with respect to experimental thermodynamic data including both solvation free energies and heats of vaporization. Only after such optimization of van der Waals parameters can we assess the qualities of both charging schemes with more certainty. This will be one of the future directions for refining existing Amber force fields in this group.

The three new united-atom van der Waals parameters were then tested in a series of relative free energy calculations described in the Method section. The relative solvation free energies between the united-atom and all-atom models should be the smallest if the van der Waals parameters are at their optimal values. However, it is impossible to reach zero difference because the atomic charges are already different between the two models. From Figure 2a, we can clearly see that the relative solvation free energies of the tested 18 amino acid side-chain analogues are very small. The average is only 0.26 kcal/mol and the largest error is only 1.40 kcal/mol for

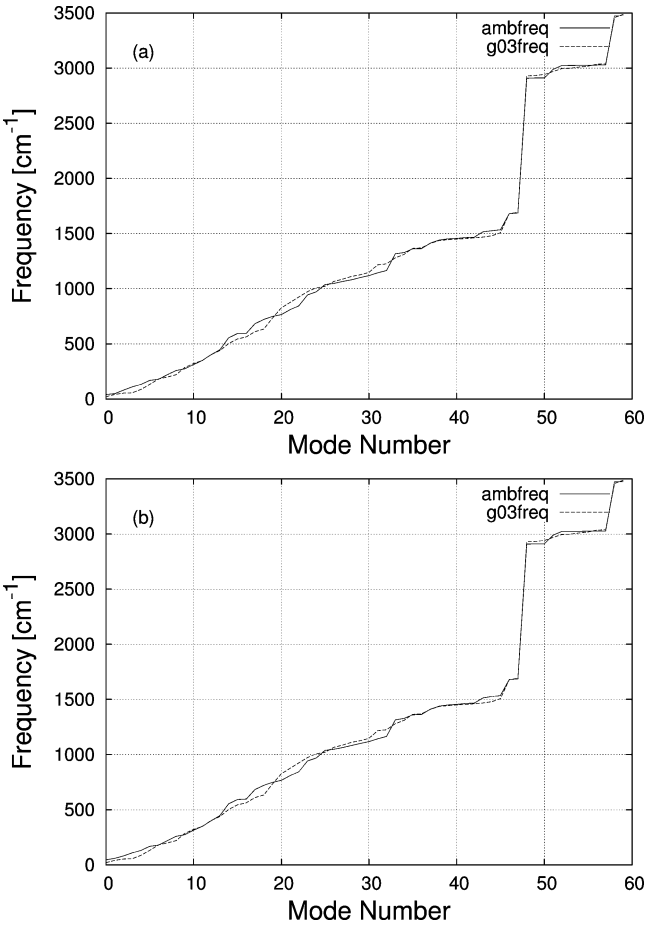

Figure 3. Gaussian03 and Amber frequencies of alanine dipeptide using the parm99 covalent terms: (a) charges are taken from the Cornell et al. all-atom force field; (b) charges are taken from the Duan et al. all-atom force field.

ASP with a rather small relative error due to its large absolute solvation free energy. As a comparison, we also calculated relative solvation free energies by perturbing the tested amino acids side-chain analogues from the Cornell et al. all-atom force field to the united-atom force field. The departure from the Cornell et al. force field is noticeable (average is 0.63 kcal/mol, the largest one is 1.88 kcal/mol, see Figure 2b). It is not surprising because we have used a different quantum mechanical model to fit atomic charges. A similar discrepancy was also observed when perturbing the Cornell et al. force field to the Duan et al. force field (average is 0.58 kcal/mol) (Figure 2c). These data clearly show that solvation free energy differences among the different force fields are mainly induced by different atomic charging schemes.

3.2. Compatibility of New Charging Scheme with Existing Covalent Terms. As mentioned in the Method section, all bond and bond angle parameters are retained from the Amber all-atom force field parm99 data set.[27] To show that the Duan et al. charging scheme (based on B3LYP/cc-pVTZ//HF/6-31g**) is compatible with the covalent parameters in parm99, normal-mode analysis was performed for the ALA dipeptide. Vibrational frequencies calculated by Gaussian03 using B3LYP/6-311+g** after scaling with a factor of 0.989 (obtained by comparing with the frequencies computed with B3LYP/6-311+g** and those from experiment)[40] were used as a reference. Figure 3a is a comparison between the Cornell et al. charging scheme/parm99 and Gaussian03, while Figure 3b is a comparison between the Duan et al. charging scheme/parm99 and Gaussian03. The rms

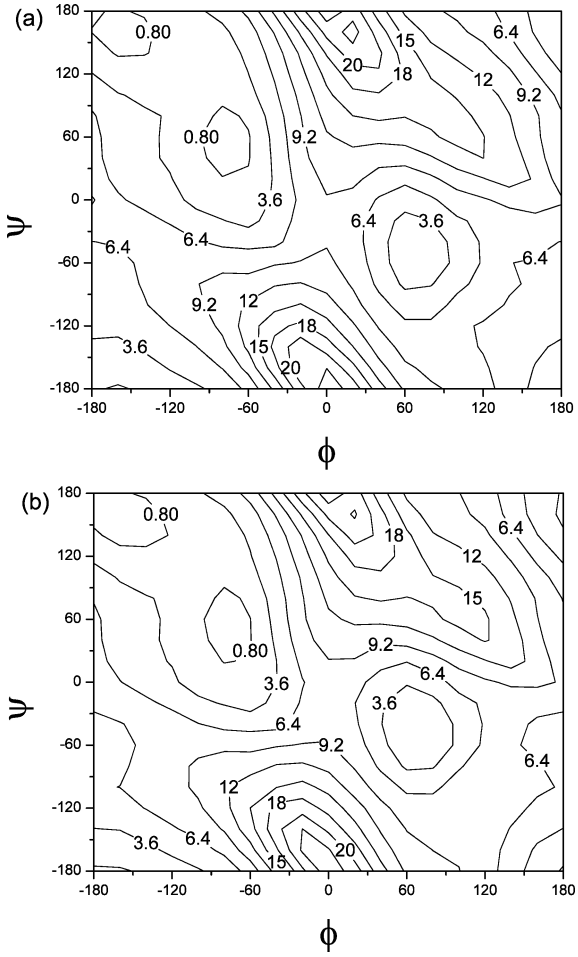

Figure 4. Main-chain relative conformational energy surfaces (with respect to $(\Phi,\Psi) = (-80.0, 80.0)$) for the ALA dipeptide calculated with (a) the Duan et al. all-atom force field and (b) the Yang et al. united-atom force field (this work).

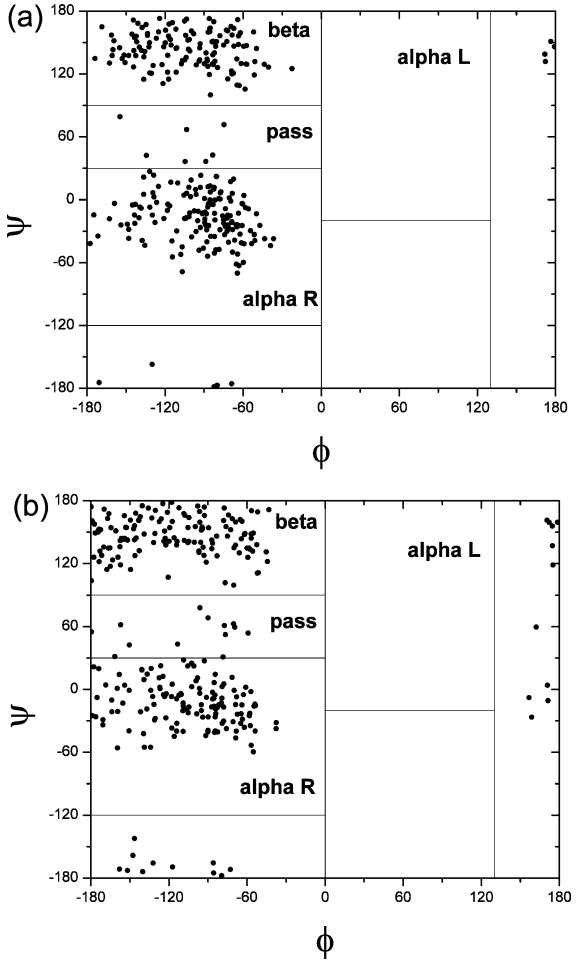

Figure 5. Main-chain torsion distributions of ILE dipeptide solvated in TIP3P at 450 K with (a) the Duan et al. all-atom force field and (b) the Yang et al. united-atom force field (this work).

deviations of parts a and b are quite similar, 66.50 and 66.40, respectively, that is, the Duan et al. charging scheme results in a similar agreement to ab initio vibration frequency for the tested system. This indicates that the Duan et al. charging scheme is compatible with parm99 covalent terms at least for the tested system.

3.3. Backbone Torsion Parameters. The backbone torsion parameters for C−N−Cα−C, N−Cα−C−N, C−N−Cα−Cβ, and N−C−Cα−Cβ were taken from the Duan et al. force field without change. To show that the Duan et al. backbone torsion parameters are applicable to the new united-atom force field, we compared the φ/ψ conformational energy surfaces of the ALA dipeptide as calculated by the all-atom force field and by the new united-atom force field. These surfaces are shown in Figure 4. The overall agreement between the all-atom and united-atom φ/ψ conformational energy surfaces is apparent. The rms deviation between the two energy surfaces is only 0.59 kcal/mol. In the low-energy regions, the two energy surfaces are almost identical, with an rms deviation of 0.32 kcal/mol. The largest energy difference between the all-atom model and the united-atom model is 1.60 kcal/mol at ($\varphi = -40°$, $\psi = 120°$). The large difference is mainly due to 1−4 van der Waals and 1−4 electrostatic interactions through energy component analysis, with a 1−4 van der Waals difference of 0.43 kcal/mol and 1−4 electrostatic difference of 1.61 kcal/mol. These results indicate that the all-atom force field backbone torsion parameters can be used in our united-atom force field at least for the simple alanine dipeptide.

TABLE 3: Percentage of Main-chain Distributions of ILE Dipeptide Solvated in TIP3P at 450 K with the Duan et al. All-atom Force Field and the Yang et al. United-atom Force Field (this work)

conformation	Duan et al.	Yang et al.
beta	44	46
pass	2	4
alpha R	51	43
alpha L	0	0
other	3	7

To further test the influence of the united-atom approximation on backbone conformational distribution, we studied a more challenging test case: ILE dipeptide with one C1, one C2, and two C3 atoms on its side chain. To make the comparison more relevant to biomolecular simulations, we performed a molecular dynamics simulation of solvated ILE dipeptide in TIP3P water molecules at 450 K for 4 ns. Comparison of the φ/ψ distribution in this simulation also shows a good agreement between the united-atom force field and the Duan et al. all-atom force field (see Figure 5 and Table 3). This further shows that we can share the all-atom force field's backbone torsion parameters in our united-atom model with a reasonable accuracy.

3.4. Side-Chain Torsion Parameters. In molecular mechanics, torsion interaction is a major determinant of relative conformational energies of a molecule. Typically, torsion terms are adjusted based on high-level ab initio conformational

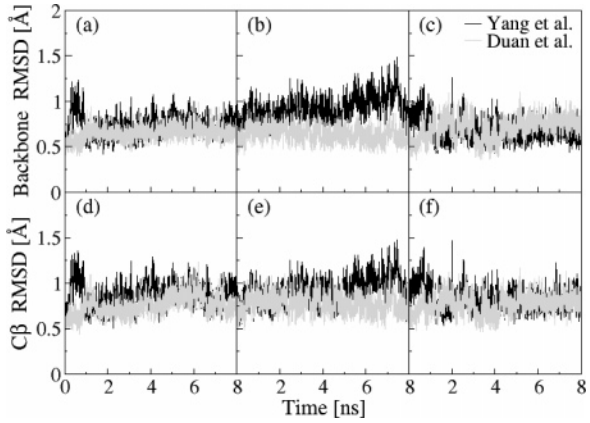

Figure 6. rms deviations of backbone heavy atoms and C_β atoms in simulations with the Yang et al. united-atom force field (this work) and the Duan et al. all-atom force field. (a), (b), and (c): Backbone heavy atoms of 1enh, 1shg, and 1pgb, respectively; (d), (e), and (f): C_β atoms of 1enh, 1shg, and 1pgb, respectively.

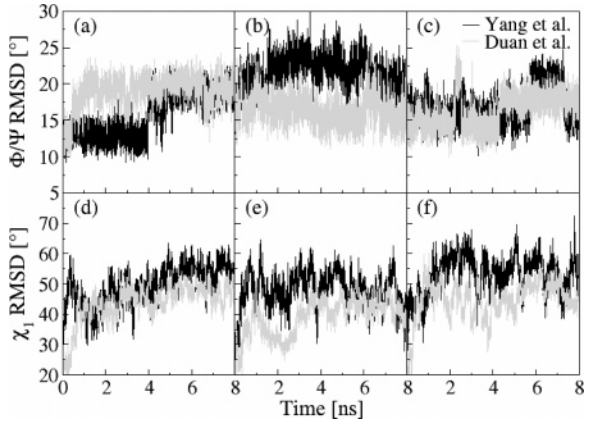

Figure 7. rms deviations of Φ/Ψ and $\chi 1$ in simulations with the Yang et al. united-atom force field (this work) and the Duan et al. all-atom force field. (a), (b), and (c): Φ/Ψ of 1enh, 1shg, and 1pgb, respectively; (d), (e), and (f): $\chi 1$ of 1enh, 1shg, and 1pgb, respectively.

energies, for example, in the development of the previous Amber force field parm99 parameter set. Here, the strategy is to optimize the torsion terms in the new united-atom force field to give the best agreement to the all-atom conformational energies that have already been optimized with respect to high-level ab initio conformational energies. After optimization, rms deviations for χ_1, χ_2, χ_3, χ_4, and χ_5 in all amino acid side chains are 0.10, 0.31, 0.44, 1.08, and 0.04 kcal/mol, respectively, for all-natural amino acid side chains found in the Amber force field database. The relative conformational energy surfaces of four side-chain torsion angles with the largest rms deviations are also shown in Figure S-1 in the Supporting Information. Table S-3 in the Supporting Information lists the optimized torsion parameters.

3.5. Simulations of Globular Proteins in Explicit and Implicit Solvents. Even if our motivation for developing the new united-atom force field is for implicit solvents, the accuracy of the new force field is presently assessed through analyzing MD simulations in explicit solvent with PME treatment of long-range interactions. This is due to the limitations of existing implicit solvents, especially in the treatment of nonpolar solvation interactions, which would make the accuracy analysis of the new force field complicated. In the following, snapshots for the three tested small globular proteins in TIP3P solvent were used to evaluate the accuracy and stability of the new force field. As a reference, simulations of the three proteins using the Duan et al. all-atom force field were also analyzed under the same conditions.

Time evolutions of rms deviations of backbone heavy atoms and side-chain Cβ atoms, respectively, were calculated for snapshots for all 10 trajectories with respect to the crystal structure. These results were then compared to those obtained with the Duan et al. force field (see Figure 6, showing the first trajectory in each set of simulations). Time evolutions of rms deviations of backbone torsion angles φ/ψ and side-chain torsion angle χ_1, respectively, along the trajectories with respect to the crystal structure were also calculated and shown in Figure 7 (showing the first trajectory in each set of simulations). In general, Time evolutions of rms deviations in the united-atom simulations are qualitatively similar to those in the all-atom simulations. The simulation data indicate that the performance of the new united-atom force field is similar to that of the Duan et al. all-atom force field for the three tested systems.

Mean simulated structures for all 10 trajectories were also computed for the three proteins. These are superimposed with

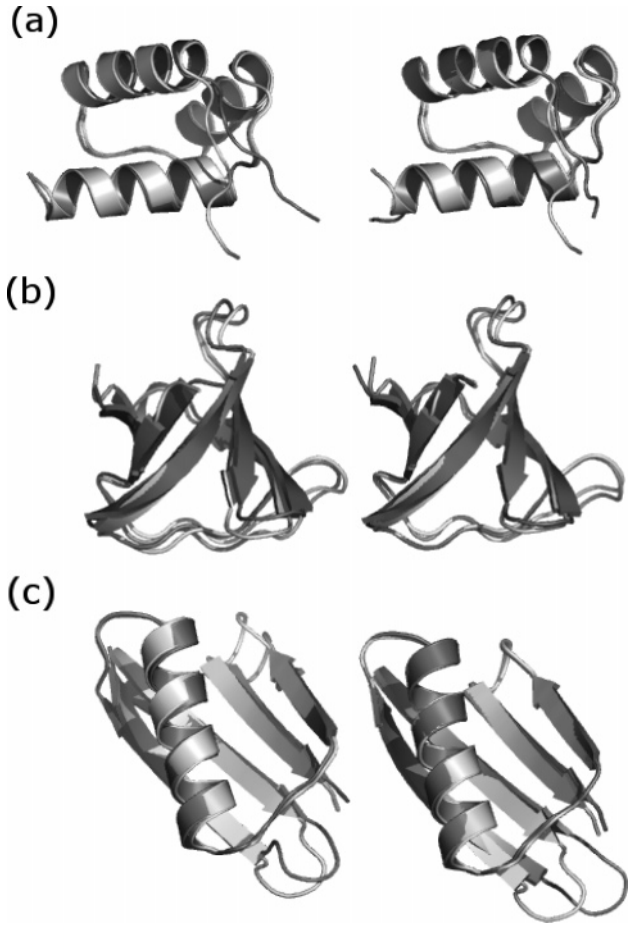

Figure 8. Mean simulated structures superimposed with crystal structures of (a) 1enh, (b) 1shg, and (c) 1pgb. Gray denotes crystal structure and black denotes mean simulated structure of the Yang et al. united-atom force field (this work, left) or the Duan et al. all-atom force field (right).

crystal structures in Figure 8. Noticeable differences in backbone conformations between the united-atom and all-atom force fields are only in the loop and/or variable regions of the three proteins. Backbone and Cβ rms deviations between the mean simulated structures and crystal structures are both computed and listed in Table 4. The rms deviations are comparable between the all-atom and united-atom models. rms deviations of backbone torsion angles φ/ψ and side-chain torsion angles χ_1 were also computed for the two mean simulated structures. Differences

TABLE 4: rms Deviations of Mean Simulated Structures Solvated in TIP3P with the Duan et al. All-atom Force Field and the Yang et al. United-atom Force Field (this work) for 1enh, 1shg, and 1pgb

	root mean squared deviations	
protein	Duan et al.	Yang et al.
	Backbone Atoms (Å)	
	0.39	0.41
	C_β (Å)	
1enh	0.50	0.54
	ϕ/ψ (deg)	
	10.21/15.09	8.44/7.32
	χ_1 (deg)	
	35.25	36.06
	Backbone Atoms (Å)	
	0.53	0.76
	C_β (Å)	
1shg	0.63	0.91
	ϕ/ψ (deg)	
	12.76/14.69	14.75/14.06
	χ_1 (deg)	
	31.40	43.36
	Backbone Atoms (Å)	
	0.47	0.63
	C_β (Å)	
1pgb	0.51	0.72
	ϕ/ψ (deg)	
	7.97/9.50	10.32/16.01
	χ_1 (deg)	
	41.16	46.94

in the φ/ψ rms deviations between all-atom and united-atom models are very small, less than 8°, consistent with our design effort to minimize the perturbation to the backbone groups. Differences in the χ_1 rms deviations between all-atom and united-atom models are larger due to the approximation of side-chain atoms in the united-atom force field. However, the differences are still reasonably small (less than 13°).

Finally, Ramanchandran plots of ALA and non-ALA amino acids excluding PRO and GLY from all 10 trajectories are also analyzed for both the united-atom and all-atom simulations. These are shown in Figures 9 and 10, respectively. Again, similar distributions between the united-atom and all-atom simulations are observed. These further show that our strategy to minimize the backbone perturbations works reasonably well in typical protein environments.

To demonstrate the per-step efficiency of the new united-atom force field in implicit solvents, short LD simulations (1000 steps) in the GB implicit solvent were also performed for the same three small proteins. The sizes of the systems and the CPU times for both the united-atom and all-atom force fields are listed in Table 5. These preliminary tests show that over 50% less hydrogen atoms are needed for these small proteins with the new united-atom force field. A speed up of about two can be observed per-step for each of the three tested small proteins.

3.6. Ab Initio Folding Simulations. To demonstrate the additional efficiency gain of the new united-atom force field for conformational sampling, we performed ab initio folding simulation of the ALA18 peptide in distance-dependent dielectric. Polyalanine peptides in distance-dependent dielectric are known toy systems with a well-defined and stable helical structure. In all 10 independent folding trajectories starting from the fully extended structure in the new united-atom force field, ALA18 is able to fold into the helical structure with first-pass folding times from 0.l to 63.6 ns (according to backbone rms deviations and relative total potential energy with respect to the equilibrium simulations of the folded helical structure, see Figure 11). The median first-pass folding time is 0.45 ns. In

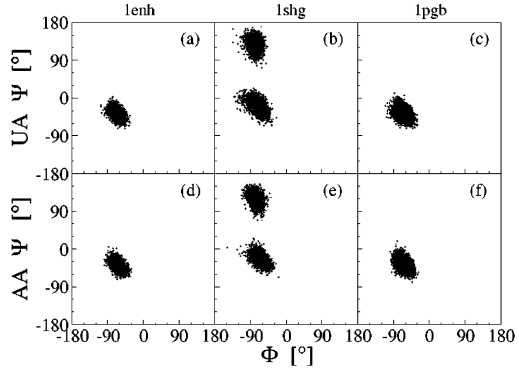

Figure 9. Main-chain torsion distributions of ALA in simulations with the Yang et al. united-atom (UA) force field (this work, (a), (b), and (c)) and with the Duan et al. all-atom (AA) force field ((d), (e), and (f)). Here, (a) and (d) are in simulations of 1enh; (b) and (e) are in those of 1shg; and (c) and (f) are in those of 1pgb.

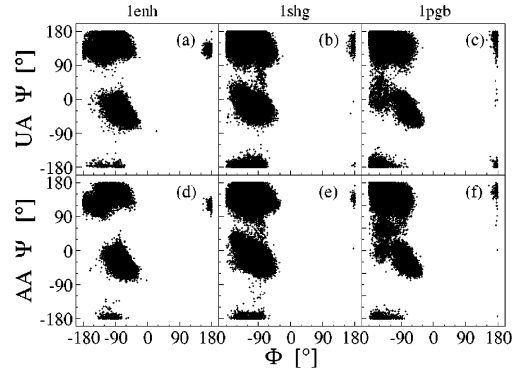

Figure 10. Main-chain torsion distributions of amino acids other than ALA, PRO, and GLY in simulations with the Yang et al. united-atom (UA) force field (this work, (a), (b), and (c)) and with the Duan et al. all-atom (AA) force field ((d), (e), and (f)). Here, (a) and (d) are in simulations of 1enh; (b) and (e) are in those of 1shg; and (c) and (f) are in those of 1pgb.

TABLE 5: Sizes and Timings of the Three Implicit Solvent Simulations with the Duan et al. All-atom Force Field and the Yang et al. United-atom Force Field (this work)[a]

	Duan et al.		Yang et al.	
protein	no. atoms/ no. H atoms	CPU time(s)	no. atoms/ no. H atoms	CPU time(s)
1enh	947/480	289.6	684/217	152.4
1shg	955/483	295.5	669/197	145.9
1pgb	855/419	238.5	622/186	125.9

[a] Timings are for 1000 steps of LD at 300 K.

contrast, ALA18 in the Duan et al. all-atom force field can only fold into the helical structure in 3 out of 50 independent folding trajectories starting from the same fully extended structure. Here, 50 trajectories were run because no folding was observed in the first 10 trajectories. The first-pass folding times are in the range from 8.7 ns to higher than 100 ns simulated for each independent trajectory. The median first-pass folding time is apparently beyond 100 ns. Even without a quantitative estimation of simulated folding rates in both force fields, which usually requires over 100 folded trajectories to fit a first-order (two-state) folding kinetics curve with reasonable accuracy, we can still obtain a conservative estimation that the new united-atom force field is a factor of 200 more efficient than the Duan et al. all-atom force field using the median first-pass folding time of ALA18 as a benchmark. Note that the gain in sampling efficiency is in addition to the per-step acceleration over the Duan et al. all-atom force field. This is very promising

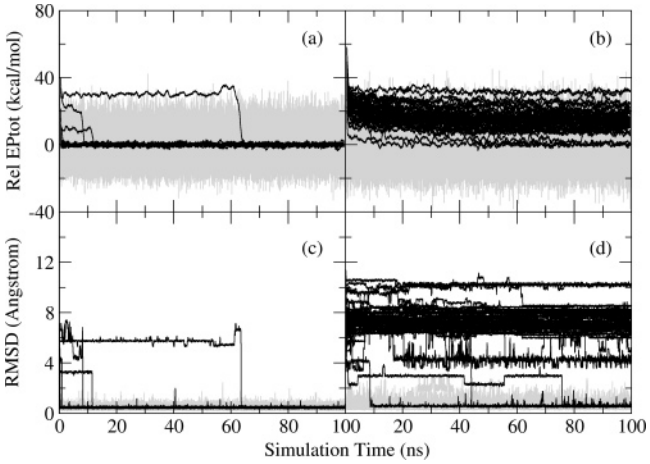

Figure 11. Ab initio folding simulations of the ALA18 peptide at 300 K with the Yang et al. united-atom (UA) force field (this work, (a) and (c)) and with the Duan et al. all-atom (AA) force field ((b) and (d)). Here, (a) and (b) are plots of relative potential energy (Rel EPtot, with respect to the average potential energy of all 10 equilibrium trajectories from the folded helical structure) vs simulation time; (c) and (d) are plots of the main-chain heavy-atom rms deviations (rmsd, with respect to the average structure of all 10 equilibrium trajectories from the folded helical structure) vs simulation time. In all figures, gray denotes instantaneous values (Rel EPtot or RMSD) vs simulation time for all 10 independent equilibrium trajectories starting from the folded helical structure, and black denotes running-averaged values (with a running window of 1000 ps to make plots clear) vs simulation time for all independent ab initio folding trajectories starting from the fully extended structure (10 for UA and 50 for AA). Note that all 10 UA ab initio folding trajectories reach the folded state within 100 ns, but only 3 out of 50 AA ab initio folding trajectories reach the folded state within 100 ns.

considering only a very modest reduction in the protein representation is adopted in the new united-atom force field. Apparently, more ab initio folding simulations of different systems will be needed to fully establish the efficiency gain of the new united-atom force field in conformational sampling.

4. Conclusion

Following the effective two-body additive formalism of all-atom Amber force fields, we have developed a new-generation Amber united-atom force field for simulations involving demanding conformational sampling. In the new united-atom force field, all hydrogen atoms on aliphatic carbons in all amino acids are united with carbons except those on Cα. Our choice of explicit representation of all backbone atoms aims at minimizing the perturbation of the united-atom approximation to the backbone conformational distributions and simplifying the development of backbone torsion terms. Tests in dipeptides and solvated proteins show that our goal of minimal perturbations of backbone conformations is achieved quite well.

The new united-atom force field uses the Duan et al. charging scheme based on B3LYP/cc-pVTZ//HF/6-31g** quantum mechanical calculations with a continuum solvent to mimic the low dielectric environment of the protein interior. van der Waals parameters of newly defined united carbons were reoptimized starting from published values. The suitability of mixing new charges and van der Waals parameters with existing Amber covalent terms was tested on alanine dipeptide and was found to be reasonable. Parameters for all new torsion terms on the side chains were also refitted based on the new charges and van der Waals parameters.

Finally, molecular dynamics simulations of three small globular proteins in the explicit TIP3P solvent were performed to test the overall accuracy of the new united-atom force field. Similar accuracies with respect to the crystal structures were observed for the new united-atom and Duan et al. all-atom force fields. The per-step efficiency of the new united-atom force field over the Duan et al. all-atom force field was also demonstrated when it was used for dynamics simulations in the implicit GB solvent. A speed up of around two per step was observed for the three tested small proteins. The efficiency gain of the new united-atom force field in conformational sampling was further demonstrated with a well-known toy protein folding system, ALA18 in distance-dependent dielectric, which is known to fold into a stable helical structure. It was found that the new united-atom force field is at least a factor of 200 more efficient than the Duan et al. all-atom force field for the tested system.

It is instructive to discuss future directions for this project. In this study, we have adopted the Duan et al. charging scheme, that is, RESP charge fitting based on B3LYP/cc-pVTZ//HF/6-31g** quantum mechanical electrostatic potentials in a continuum solvent. In this way, one of the limitations of previous Amber force fields—lacking a controlled representation of polarization effect in the condensed phase—can be partially resolved, though still not in a perfect fashion. Indeed, it can be argued that a proper polarization is probably not feasible without explicit representation of polarization in a force field. The Duan et al. charging scheme intends to achieve such a balance by using a low dielectric similar to the protein interior in quantum mechanical calculations. However, use of the protein interior dielectric may not in general guarantee an optimal balance between solvation and desolvation of all amino acids. A validation and optimization process against liquid-state thermodynamic properties as in the development of OPLS force field will need to be followed to further optimize the low dielectric value in quantum mechanical calculations for the charge-fitting process. It is likely that the "best" dielectric value to strike an "optimal" balance in polarization due to solvation and desolvation of all amino acids may not be that of the protein interior.

An equally important issue is the optimization of van der Waals parameters. Here, we have chosen to stay with the simplistic van der Waals scheme in Amber force fields. However, analysis of side-chain solvation free energies shows quite a few noticeable differences between experiment and simulation for both the Duan et al. and Cornell et al. charging schemes. Thus, more attentions should also be paid to van der Waals parameters in future Amber force field developments.

Acknowledgment. This project was originated from Dr. Peter Kollman's group at UCSF. Dr. Peter Kollman died unexpectedly in May, 2001. We dedicate the manuscript to his memory. We thank all members of the Amber development team, especially Dr. David Case, for valuable inputs and advice for this project. This project is supported by the state of California and NIH (GM069620).

Supporting Information Available: Table S-1: United carbon atoms and their bonded hydrogen atoms in all-natural amino acids in the new united-atom force field. Table S-2: Atomic point charges of natural amino acids in the new united-atom force field. Table S-3: Side-chain torsion parameters related to the new united carbon atom types in the new united-atom force field. Figure S-1: Relative conformational energy profiles of four side-chain torsion angles where largest deviations between the new united-atom and Duan et al. all-atom force fields are observed during fitting. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

(2) 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.

(3) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(4) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.

(5) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(6) Taketomi, H.; Ueda, Y.; Go, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445.

(7) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.

(8) Eyrich, V. A.; Standley, D. M.; Friesner, R. A. *J. Mol. Biol.* **1999**, *288*, 725.

(9) Moult, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 194.

(10) Dunfield, L. G.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2609.

(11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(12) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.

(13) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244.

(14) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035.

(15) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383.

(16) Lwin, T. Z.; Zhou, R. H.; Luo, R. *J. Chem. Phys.*, in press.

(17) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(18) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.

(19) Jorgensen, W. L.; Tiradorives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(20) Tiradorives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 2773.

(21) Orozco, M.; Tiradorives, J.; Jorgensen, W. L. *Biochemistry* **1993**, *32*, 12864.

(22) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

(24) 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. *J. Comput. Chem.* **2003**, *24*, 1999.

(25) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(26) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(27) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.

(28) Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535.

(29) Johnson, R. D. I. NIST Computational Chemistry Comparison and Benchmark Database. In *NIST Standard Reference Database*; NIST: Gaithersburg, MD, 2005; Vol. 101.

(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(31) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; San Francisco, CA, 2004.

(32) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tiradorives, J. *J. Chem. Phys.* **1988**, *89*, 3742.

(33) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(34) Hockney, R. W.; Eastwood, J. W. *Computer Simulations Using Particles*; McGraw-Hill: New York, 1981.

(35) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(36) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1989.

(37) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(38) Lwin, T. Z.; Zhou, R. H.; Luo, R. *J. Chem. Phys.* **2006**, *124*.

(39) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.

(40) Foresman, J. B.; Frisch, A. *Exploring Chemistry With Electronic Structure Methods: A Guide to Using Gaussian*, 2nd ed.; Gaussian, INC.: Wallingford, CT, 1996.