JCTC Journal of Chemical Theory and Computation

Toward a Coarse-Grained Protein Model Coupled with a Coarse-Grained Solvent Model: Solvation Free Energies of Amino Acid Side Chains

Wei Han,[†] Cheuk-Kin Wan,[†] and Yun-Dong Wu*[,†,‡]

Department of Chemistry, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, China, and National Laboratory of Molecular Sciences, College of Chemistry, Peking University, Beijing, China

Received May 23, 2008

Abstract: Recently, we reported that molecular dynamics (MD) simulations using a coarse-grained (CG) peptide model coupled with a CG water model are able to reproduce many of the structural and thermodynamic features of short peptides with nonpolar side chains at 10^3 times the normal speed (*JCTC*, **2007**, *3*, 2146−2161). To further develop a CG protein model for MD simulations, we systematically parametrized the side chains of all 20 naturally occurring amino acids. We developed the parameters by fitting the dihedral potentials of 13 small molecules, the densities and self-solvation free energies of liquids of eight organic molecules, and the hydration free energies of 35 small organic molecules. In a set of 11 classes of compounds (105 in total) including alkanes, alcohols, ethers, ketones/aldehydes, amines, amides, aromatics, carboxylic acids, sulfides/thiols, alkyl ammoniums, and carboxylate ions, the average error in the calculated hydration free energies compared with experimental results is about 1.4 kJ/mol. The average error in the calculated transfer free energies of the 19 side-chain analogues of amino acids from cyclohexane to water is about 2.2 kJ/mol. These results are comparable to the results of all-atom models.

1. Introduction

Coarse-grained (CG) force fields have become promising tools for studies of protein folding and protein−protein interactions.[1−12] Unlike the widely used all-atom force field, CG force fields represent a group of atoms with a single particle. The simplicity of the model and smoother potential energy surfaces make CG force fields fast enough to simulate protein folding at biologically relevant time scales, which is currently difficult for all-atom force field models. There have been many applications of CG force fields to protein simulations at various levels. CG models, which explicitly represent each heavy atom, have been used to study the folding of small peptides with about 20 amino acids, such as polyalanine and Trp-cage.[1,2] CG models with coarser resolution have been used to study the formation of helix bundles and the aggregation of peptides.[3−5] A residue-based CG model is even feasible for the study of the stability and dynamics of a viral capsid.[6] Moreover, Voth and co-workers have recently developed a systematical way to derive interaction parameters between different scales by matching CG interaction force with the simulated force in all-atom simulation.[13] This method allows the CG model to study a multiscale system with solvent or membrane environments as coarse-grained model but with proteins represented in more detail.[14−16]

The quality of CG simulations of proteins relies on the interparticle interaction parameters, particularly for hydrogen bonds (H-bonds) and hydrophobic interactions. These interactions are known to depend strongly on the local environment, such as the solvation level, and to have many-body characters.[10,17] They are pairwise additive in many CG force fields. So far, there has not been much effort in parametrize CG force fields with solvation properties except for a recent CG force field for the simulations of biomolecules in lipid/

* Corresponding author e-mail: chydwu@ust.hk.
[†] The Hong Kong University of Science & Technology.
[‡] Peking University.

10.1021/ct800184c CCC: $40.75 © 2008 American Chemical Society
Published on Web 10/15/2008

water environment, namely the MARTINI force field, which was developed by Marrink and co-workers.[18−20] The way of parametrization in this model has been recognized to be critical in reproducing the thermodynamic properties of systems in a comparative study with all-atom simulation.[21] Their CG force field has been successfully used to study the formation and fusion of micelles and the phase transitions of membranes.[22−25] By including a residue-based protein model, the MARTINI protein model can reproduce the transfer free energies of 16 amino acids side-chain analogues from water to the center of lipid bilayer with an average error of about 8 kJ/mol with respect to all-atom simulations.[19] This model allows the CG model to study interactions between proteins and lipid membrane and dynamics and thermodynamics of proteins in membranes beyond microsecond time scale.[26−32]

Biological processes such as protein folding, protein−protein recognition, and aggregation of transmembrane helices are all associated with the transfer of amino acids from a polar environment such as water to a nonpolar environment.[33−36] Thus, solvation properties of amino acids, particularly the solvation changes in different media, are very important in controlling these processes.[37−41] Indeed, experimental solvation properties of various organic compounds have been used to parametrize several all-atom force fields such as OPLS and GROMOS.[42−47] Recently, the solvation free energies of side-chain analogues of amino acids were studied by OPLS and GROMOS all-atom force fields simulations.[48−50] These simulations reproduced transfer free energies of side-chain analogues of amino acids from cyclohexane to water determined experimentally with reasonable accuracy (2.4−4.2 kcal/mol).

We recently reported our initial effort in developing a CG protein force field model[51] that couples with the CG solvent model developed by Marrink et al.[18] Our model represents amino acid backbone by four particles to include more detail. We developed parameters for peptide backbones, aliphatic side chains, and protein-CG water interactions based on solvation free energies of small molecules. Initial application of the model to the simulation of Ac-$(Ala)_6$-X-$(Ala)_7$-NHMe, X = Ala, Gly, Val, and Leu, indicated the following: (1) this model is 100−1000 times faster than all-atom models; (2) various secondary structures such as helices, turns, and sheet structures can be generated; (3) the helix-coil transition free energies of these peptides are well reproduced; and (4) the structural information from the simulations allows the analysis of the difference in the helical propensities of different amino acids. In this paper, we report on the extension of the parametrization to all 20 naturally occurring amino acid residues. The parametrization is mainly based on solvation properties and can be divided into three parts: (1) all geometric and dihedral parameters are optimized based on conformational energies of organic compounds by quantum mechanics calculations;[46,52] (2) nonbonded interactions are then parametrized by fitting the densities and self-solvation free energies of pure organic liquids and the solvation free energies of organic compounds in water; and (3) the transfer free energy of side-chain analogues of various amino acids from cyclohexane and water were also computed and used to benchmark the performance of our model. We show that the parameters are transferable and the accuracy of the model in determining the hydration free energies of 105 small molecules is comparable to the accuracy of all-atom models.

2.1. The Coarse-Grained (CG) Protein Model. The CG protein model has been described in detail in a previous work.[51] Briefly, our protein model is essentially a united-atom model with each CG particle representing a single heavy atom together with the hydrogen atoms attached to it. The potential energy of the system includes bonded and nonbonded terms as shown below:

$$V_{total} = V_{bonded} + V_{non-bonded} \tag{1}$$

The bonded terms (V_{bonded}) include quadratic potentials for angle bending (V_{angle}) or to conserve sp^2 planarity and sp^3 configuration ($V_{improper}$) and a combination of two types of potentials ($V_{torsion}$ and $V_{14,ij}$) to describe dihedral angles:

$$V_{Angle} = K_{Angle}(\theta - \theta_0)^2/2 \tag{2}$$

$$V_{Improper} = K_{Improper}(\xi - \xi_0)^2/2 \tag{3}$$

$$V_{torsion} = K_{torsion}(1 + \cos(n\varphi - \varphi_0)) \tag{4}$$

$$V_{14,ij} = \sum_{1-4relationship} 4\varepsilon_{14,ij}\left(\frac{\delta_{14,ij}^{12}}{r^{12}} - \frac{\delta_{14,ij}^{6}}{r^{6}}\right) \tag{5}$$

Both K_{angle} and $K_{improper}$ take the same values, 300 kJ/mol/rad^2, as in our previous work. In addition, all bonds are constrained at their equilibrium length (r_0) by the LINCS algorithm.[53] Finally, all the nonbonded terms are expressed as Lennard-Jones (LJ) potentials:

$$V_{non-bonded} = \sum_{i<j} 4\varepsilon_{ij}\left(\frac{\delta_{ij}^{12}}{r^{12}} - \frac{\delta_{ij}^{6}}{r^{6}}\right) \tag{6}$$

2.2. Simulation Setup. The simulations were performed with the GROMACS 3.3.1 package.[54] For all simulations, the van der Waals (vdW) interactions have a cutoff of 1.2 nm, and they were smoothed to zero from 0.9 to 1.2 nm. The temperature and pressure are controlled by a thermostat and a pressure bath, with coupling constants of 0.1 and 0.5 ps, respectively.[55] The time interval to integrate the Newton equations is 10 fs, and the neighboring list is updated every 10 steps. In the preparation stage, the whole system is subjected to 5000 steps of steep descent optimization and then to a 1200 ps of pre-equilibrium at 300 K and 1 atm. The generated coordinates are used for free energy calculations.

2.3. Solvation Free Energy Calculations. The solvation free energy, ΔG_{sov}, is defined as the free energy difference between the state where the solute is immersed in the solvent and the state where the solute is isolated from the solvent. The free energy perturbation method calculates the solvation free energy by introducing a coupling parameter (λ) with interaction potentials between the solute and the solvent. As λ gradually varies from zero to unity, the interactions are gradually turned on. During this process, ΔG_{sov} can be calculated as[56]

$$\Delta G_{sov} = G_1 - G_0 = \int_{\lambda=0}^{\lambda=1} d\lambda \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \tag{7}$$

Table 1. Summary of Uncertainties of Calculated Solvation Free Energies of Various Solution Systems

solution systems	CG particle number	total time of perturbation (ns)	uncertainties (kJ/mol)
pure liquids[a]	1000−2000	128	<1
water solutions[b]	350	300	0.2−1.5
cyclohexane solutions[c]	1250	128	<1

[a] Pure liquids of cyclohexane, n-pentane, isopentane, 2,3-dimethyl-2-butene, diethyl ether, triethylamine, benzene, and dimethylsulfide were used to derive their self-solvation free energies. [b] The solutes include 105 organic molecules. [c] The solutes include 19 side-chain analogues of amino acids.

where $U(\lambda)$ is the total energy when the system is in the intermediate state, λ. For each perturbation calculation, there are 32 intermediates (λ=0.0, 0.02, 0.06, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.44, 0.47, 0.5, 0.53, 0.56, 0.59, 0.62, 0.65, 0.68, 0.71, 0.73, 0.75, 0.78, 0.81, 0.84, 0.87, 0.9, 0.92, 0.94, 0.96, 0.98, 1.0). In addition, a soft-core Lennard-Jones potential is applied to avoid the singularity problem when λ is close to unity or zero.[57]

We carried out free energy calculations for several solution systems including pure liquids of eight organic compounds, aqueous solutions of 105 organic compounds, and cyclohexane solutions of 19 side-chain analogues of amino acids. The detailed information is given in Table 1. The uncertainties of the calculations are estimated by performing three calculations under the same conditions with different initial atomic velocities. The error is larger for water than for nonpolar solvent. It also increases with the number of internal degrees of freedom of the solutes. All the reported solvation free energies were averaged results. As shown in Table 1, the uncertainties of our calculations are less than 1 kJ/mol in most cases, except for water (0.2−1.5 kJ/mol). The calculated uncertainties are comparable to those from similar solvation free energy calculations using all-atom force fields reported by Chang et al. (0.4−1.2 kJ/mol)[50] and MacCallum and Tieleman (less than 2.5 kJ/mol).[48]

3. Results and Discussion

There are 24 types of protein particles in our current model used to represent various kinds of amino acids (Table 2). Each of them represents a single heavy atom or a single heavy atom with its attached hydrogen atoms. To couple our model with the CG solvent model developed by Marrink et al.,[18] we also included the CG water particle type, which includes four water molecules.

Therefore, the optimization of the parameters of these particle types was divided into two steps. First, the interaction parameters among protein particle types were optimized. Then the interactions between the CG water particles and protein particles were parametrized. The interaction parameters between the CG water particles were taken directly from the reported results by Marrink et al.[18]

The parametrization of the interactions among the protein particles was further divided into three steps: (1) the equilibrium bond length, r_0, and angle value, θ_0, were obtained from the optimized geometries of organic molecules with quantum calculations; (2) the dihedral terms (ϵ_{14}, δ_{14}, K_{torsion}, n, and ϕ_0) were then optimized by fitting the conformational energies of the organic molecules with quantum calculations; and (3) the nonbonded terms (ϵ and δ) were optimized by fitting the liquid densities and self-solvation free energies of several organic compounds.

Table 2. Summary of Particle Types

type	description
CH_4	sp^3 carbon in methane
-CH_3	sp^3 carbon with three H
-CH_2-	sp^3 carbon with two H
>CH-	sp^3 carbon with one H
-CH=	aromatic CH group
>C=O	carbonyl carbon
-COO^-	carboxylate carbon
-CH_x-P	sp^3 carbon directly connected to polar groups (P)
>N-	sp^3 nitrogen without H
>NH	sp^3 nitrogen with one H
-NH_2	sp^3 nitrogen with two H
-CO-**N<**	N-disubstituted amide
-CO-**NH**-	N-monosubstituted amide or heterocycle NH
-CO-**NH_2**	amide NH_2
=NH	NH doubly bonded to C
=N-	heterocycle N
-NH_3^+	ammonium NH_3^+
-OH	hydroxyl group
-CO-**OH**	OH group of carboxylic acid
-O-	ether oxygen
>C=**O**	carbonyl oxygen
-C**OO**$^-$	carboxylate oxygen
-SH	thiol SH
-S-	sulfide S
W	CG water

3.1. Optimization of Geometric Parameters. All the geometric parameters including bond lengths, r_0, and angle values, θ_0, were obtained by the optimization of 15 small organic molecules via quantum mechanics calculations with the GAUSSIAN 03 program package.[58] All the calculations were performed at the RHF/6−31G* level. These small molecules and the derived parameters are given in Figure S1 in the Supporting Information (SI).

3.2. Optimization of Dihedral Parameters. Thirteen organic molecules were used to derive the dihedral terms including ϵ_{14}, δ_{14}, K_{torsion}, n, and ϕ_0, while the optimized parameters are given in Table 3. The simulated conformational energies and the calculated data by Jorgensen et al. (RHF/6−31G*)[46] and by Halgren[52] (MP4SDQ/TZP)[59,60] are given in Table 4. The conformers of the small molecules can be divided into two categories: stable minima, such as *gauche* (±60°) and *trans* (±180°) conformers of n-butane, and transition states, such as *cis* (0°) and *skew* (±120°) conformers of n-butane. The average energy deviation of the stable minima in our calculations from those by Jorgensen et al. was about 1.4 kJ/mol, mainly resulting from the large deviations of *gauche-trans* relative energies in n-butane, methyl ethyl ether, and propanethiol, which were about 2.0, 3.7, and 1.8 kJ/mol, respectively. However, when compared with the data from the MP4SDQ/TZP calculations, the above three deviations dropped to 0.7, 2.6, and 0.3 kJ/mol, respectively. Also, the deviations for n-butane and methyl ethyl ether from experimental values were 0.6 and 2.9 kJ/mol, respectively.[61,62] Compared with the energies of the transition states from Jorgensen et al., the deviation in our results was about 5.0 kJ/mol. The barriers calculated from our model were systematically lower than those from the

Table 3. Parameters for $\epsilon_{14,ij}$, $\delta_{14,ij}$, ϕ_0, $K_{torsion}$, and n

pair type	$\epsilon_{14,ij}$ (kJ/mol)	$\delta_{14,ij}$ (nm)
C_{sp3}—C_{sp3}[a]	0.1	0.360
-**O**—/—**OH**/-CO-**OH**—C_{sp3}	0.6	0.300
>C=**O**/-C**OO**$^-$—C_{sp3}	0.6	0.280
N[b]—C_{sp3}	0.1	0.345
>**C**=O/-C**OO**$^-$/>**C**<—C_{sp3}	0.1	0.320
-**S**—/—**SH**—C_{sp3}	0.1	0.380

torsion type	ϕ_0(°)	K_{tor} (kJ/mol)	N
Y[c]-C_{sp3}-C_{sp3}/O_{sp3}/N_{sp3}[d]-Y	0	4.9	3
Y—S—C_{sp3}-Y	0	3.9	3
C_{sp3}-C_{sp3}-N_{sp2}-(C=NH)	0	3.0	3
C_{sp3}-C_{sp3}-(CO)-O^-/C_{sp3}	180	2.8	2
C_{sp3}-C_{sp3}-(CO)-NH/NH_2	180	0.8	2
C_{sp3}-N_{sp2}-(CO)-C_{sp3}	180	42.0	2
C_{sp3}-N_{sp2}-(C=NH)-NH_2	180	42.0	2

[a] C_{sp3} indicates -CH_3, >CH_2, >CH-, and -**CH$_x$**-P. [b] N indicates all types of nitrogen containing groups. [c] Y indicates any type of particles. [d] O_{sp3} and N_{sp3} indicate all types of sp^3 oxygen and nitrogen, respectively.

RHF/6−31G* calculations, indicating that it may be easier to change the conformations of the molecules in our model. However, this may have little effect on the equilibrium conformational properties. Alternatively, the RHF/6−31G* calculations may overestimate the barriers. Halgren has pointed out the inaccuracy of RHF/6−31G* in calculating the energies of conformers.[52] For example, the experimental value of the *cis-trans* energy difference of n-butane was actually about 20.4 kJ/mol,[62] lower than that (25.9) from the RHF calculations but closer to ours (17.7). However, the barrier heights of various small molecules still need to be estimated by quantum mechanics calculations at a higher level such as by MP4SDQ/TZP to further refine our dihedral potentials.

3.3. Optimization of Nonbonded Interactions among Protein Particles. Our CG model separately treats electrostatic interactions between protein particles such as hydrogen bonds by using other special interaction potentials, as described in our previous work.[51] Thus, our current task is only to obtain van der Waals (vdW) parameters (ϵ and δ) for interactions between protein particles (eq 6). Adopting the method used in the parametrization of OPLS all-atom force fields,[42−46] we optimized these nonbonded parameters by simulating liquid of organic compounds and fitting their experimental thermodynamic properties.

Specifically, δ captures the size of a particle and is related to the density of liquid composed of this type of particle. The ϵ captures the interaction strength between the particles, and it is connected to the solvation free energy of a particle in the solvent. Thus, it is possible to optimize ϵ and δ by fitting experimental liquid density and solvation free energies. In principle, all ϵ_{ij} and δ_{ij} between any two types of particles could be optimized. In our optimization procedure, we follow a simple approach. We first optimize the parameters between the same protein particle types (ϵ_{ii} and δ_{ii}) from the liquid densities and the self-solvation free energies of the organic compounds. Then, we derive the parameters for two different particle types (ϵ_{ij} and δ_{ij}) according to the Lorentz−Berthelot (LB) combination rule:

$$\delta_{ij} = {}^1\!/_2(\delta_{ii} + \delta_{jj}) \tag{8}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{9}$$

For vdW-type interactions, the rules are physically reasonable and mathematically convenient.[63] They have been widely used in all-atom force field models. To reduce the number of parameters to be optimized, the parameters for some protein particle types (ϵ_{ii} and/or δ_{ii}) were directly used for other protein particle types. The validity of the transfer of parameters was examined by simulating the solvation of various compounds in cyclohexane, which is discussed in Section 3.6.

The interactions for alkyl-type particles including primary (-**CH$_3$**), secondary (>**CH$_2$**), and tertiary (>**CH**-) carbon were first parametrized since alkyl groups are the main parts of various amino acid side chains. Cyclohexane, which has only CH_2 groups, was used to derive the parameters for >**CH$_2$**. With the parameters for >**CH$_2$** and applying the LB combination rule, the parameters for -**CH$_3$** were obtained from the simulations of n-pentane, which has both CH_2 and CH_3 groups. In a similar manner, the parameters for >**CH**- were obtained from the simulations of isopentane. In addition, the parameters for aromatic carbon type (=**CH**-) were obtained by simulating liquid benzene.

For particle types containing O, N, and S, we used aprotic solvents instead of protic solvents to derive their nonbonded parameters. Since we were interested only in the vdW parameters between these particles, this approach avoids the complication of electrostatic interactions such as intermolecular hydrogen bonds, which are present in protic solvents.

We chose diethyl ether for sp^3 oxygen (-**O**-). Diethyl ether is the smallest molecule appropriate for our purpose since it has a measurable liquid−gas equilibrium vapor pressure to calculate the self-SFE experimentally. As protic solvents cannot be used, the parameters for -**OH** cannot be directly derived. As an approximation, the ϵ_{ii} and δ_{ii} of the -**OH** group were considered the same as those of sp^3 oxygens, given that the vdW interaction of -**OH** is mainly from the oxygen.

Carbonyl groups and carboxylic groups that have sp^2 carbons (>**C**=O and -**C**OO$^-$) and oxygen (>C=**O** and -C**OO**$^-$) are important parts of peptides. The parameters for the carbonyl group can be derived from simulations of acetone. Nevertheless, the dipole−dipole electrostatic interaction among acetone molecules is considerable in liquid acetone since an acetone molecule has a dipole moment of 2.91 D, which should be avoided in this work. Thus, we obtained the vdW parameters for the sp^2 carbon by fitting the experimental data of 2,3-dimethyl-2-butene. For sp^2 oxygen, the δ_{ii} was taken from the crystal structure survey by Chothia and co-workers,[68] which is 0.280 nm, while the ϵ_{ii} was assumed to be the same as that of an -**O**- group.

In a similar manner, the vdW ϵ_{ii} and δ_{ii} of sp^3 nitrogen (>**N**-) were obtained through simulations of pure triethylamine (TEA). These vdW parameters were also applied to all other kinds of sp^3 and sp^2 nitrogens. This assumption may be reasonable since the statistical survey of crystal structures reveals that various kinds of sp^3 and sp^2 nitrogen groups share the same vdW radii.[68] Finally, the ϵ_{ii} and δ_{ii} of

Table 4. Relative Energies (kJ/mol) of Different Conformers of Various Organic Molecules Computed by our CG Model and the Quantum Calculations at the RHF/6-31G*46 and MP4SDQ/TZP[52] Levels

compound	dihedral	conformer (degree)	coarse grained	RHF/ 6−31G*	MP4SDQ/TZP
butane	C−C−C−C	0	17.7	*25.9*	N/A
		60	2.2	4.2	2.7
		120	9.9	15.3	N/A
		180	0	0	0
propanol	C−C−C−O	0	14.7	22.6	N/A
		60	0.4	0	0.5
		120	9.3	16.4	N/A
		180	0	0	0
ethyl methyl ether	C−C−O−C	0	21.1	28.6	N/A
		60	3.3	7	5.9
		180	0	0	N/A
propyl- ammonium ion	C−C−C−N	0	15.9	22.6	N/A
		60	1.3	2.04	0.3
		120	9.8	15.9	N/A
		180	0	0	0
propylamine	C−C−C−N	0	15.9	23.7	N/A
		60	1.6	2.3	1.4
		120	9.9	17.1	N/A
		180	0	0	0
butanone	C−C−C−O	0	0	0	0
		110	5.7	7	4.1
		180	12.4	12.2	N/A
proanamide	C−C−C−N	0	6.6	7.1	N/A
		180	0	0	N/A
butanamide	C−C−C−C(O)	0	17.9	24.2	N/A
		60	1.9	1.1	N/A
		120	10.5	12.2	N/A
		180	0	0	N/A
propanoate ion	C−C−C−O	0	0	0	N/A
		90	3.7	3	N/A
butanoate ion	C−C−C−C(O)	0	20.7	24.2	N/A
		60	1.8	0.3	N/A
		120	12.5	11.3	N/A
		180	0	0	N/A
N-propaneguanidine	C−C−N−C	0	28.6	34.8	N/A
		90	4.0	2.8	N/A
		120	5.7	5.4	N/A
		180	0	0	N/A
propanethiol	C−C−C−S	0	19.5	26.7	N/A
		60	2.1	3.9	1.8
		120	9.9	15.1	N/A
		180	0	0	0
ethyl methyl sulfide	C−C−S−C	0	13.6	18.8	N/A
		60	1	2.3	N/A
		120	7.8	7.6	N/A
		180	0	0	N/A

Table 5. Optimized Nonbonded Parameters (ϵ_{ij} and δ_{ij}) for Interactions between Small Particles

interaction type	ϵ_{ii} (kJ/mol)	δ_{ii} (nm)
>**CH$_2$**	0.45	0.390
-**CH$_3$**	1.00	0.390
>**CH**-	0.20	0.390
=**CH**-	0.45	0.375
-**O**-, -**OH,** -CO-**OH**	0.80	0.290
>**C**=O, -**C**OO$^-$	0.25	0.360
>C=**O,** -C**OO**$^-$	0.80	0.280
N[a]	1.00	0.330
-S-, -**SH**	2.50	0.340

[a] N indicates all types of nitrogen-containing groups.

Table 6. Comparison of Densities and Self-Solvation Free Energies between Experiments and Calculations

	density (g/cm^3)		self-SFE (kJ/mol)	
pure liquid	cal.[a]	exp.[b]	cal.	exp.[c]
cyclohexane	0.776	0.779	−17.8	−18.5
n-pentane	0.625	0.626	−13.6	−14.2
isopentane	0.635	0.620	−13.5	−13.5
benzene	0.848	0.879	−17.9	−19.1
diethyl ether	0.691	0.713	−14.6	−14.4
2,3-dimethyl-2-butene[d]	0.659	0.708	−20.2	−17.7
triethylamine	0.764	0.726	−18.9	−18.8
dimethylsulfide[e]	0.816	0.846	−16.0	−15.7

[a] All the calculated values are obtained at 300 K. [b] Reference 64. [c] Measured at 298.15 K.[65] [d] The experimental self-solvation free energy of 2,3-dimethyl-2-butene was converted from the experimental vapor pressure of 2,3-dimethyl-2-butene at 298 K.[66] [e] The experimental self-solvation free energy of dimethylsulfide was converted from the experimental vapor pressure of dimethylsulfide at 293 K.[67]

sulfur in thiol and sulfide were obtained through simulations of pure dimethylsulfide.

As shown in Table 6, our parameters of small-small particle interactions reproduced the densities of eight pure

liquids with an average error about 3.2%, and the self-SFEs of these molecules had an average deviation of about 0.7 kJ/mol.

3.4. Optimization of Nonbonded Parameters between Protein Particles and CG Water Particles. One possible way to optimize the interaction parameters between protein particles and CG water particles could be to generate ϵ_{ij} and δ_{ij} from ϵ_{ii} and δ_{ii} of protein particles and CG water particles according to the LB combination rules. However, the LB rules are appropriate only if ϵ_{ii} and δ_{ii} are both associated with vdW interactions. The previously derived ϵ_{ii} for CG water reflects the overall effect of both vdW and electrostatic interactions between groups of four water molecules. Therefore, the LB rule cannot be fully used to obtain the ϵ_{ij} and δ_{ij} for interactions between protein particles and CG water particles. One the other hand, it would be useful to utilize the LB rules partially to simplify the optimization procedure. Thus, we adopted an optimization procedure in which (1) we utilized the LB rule to generate most of the δ_{ij}, assuming that the size of the cavity created by a small particle in CG water is the vdW volume of the particle and (2) the ϵ_{ij} was obtained by reproducing the hydration free energies (HFE) of various classes of organic molecules. In our optimization procedure, the HFEs of 35 small molecules covering 11 classes of organic compounds, including alkanes, alcohols, ketones/aldehydes, ethers, carboxylic acids, amines, amides, aromatics, sulfides/thiols, alkyl ammoniums and carboxylate ions (Table 7), were used to determine the parameters of interactions between protein particles and CG water particles (Table 8).

Alkanes. All the δ_{ij} for interactions between alkyl particles and CG water particles were generated by the LB rules. The ϵ_{ij} for **CH_4** - W interactions was obtained by fitting the HFEs of methane. The ϵ_{ij} values for -**CH_3** -W, >**CH_2** -W, and >**CH**- -W were optimized by fitting the HFEs of ethane, propane, and isobutane simultaneously. As shown in Table 9, these parameters can accurately reproduce HFEs of the other ten alkane molecules that have linear, branched, and cyclic shapes. The average deviation of the calculated HFEs from experimental values was about 1.0 kJ/mol for these ten molecules and about 0.7 kJ/mol for all 14 alkane compounds.

Alcohols and Ethers. Methanol, ethanol, and 1-propanol were used to optimize the parameters for W and -**OH** interactions. During the optimization, we found that if the δ_{ij} of the W- -**OH** interaction was taken as 0.375 nm according to the combination rule, the magnitude of the HFEs of alcohols dropped too rapidly with an increasing length of the hydrocarbon chain. This trend does not agree with experimental data. We found that this problem can be alleviated by reducing the δ_{ij} for the W- -**OH** interaction to 0.280 nm. Our observations may be understandable since a polar -**OH** group has hydrogen bonds with individual water molecules and four water molecules in a CG water particle may not appear equivalent to an -**OH** group. Thus, the δ_{ii} for CG water, which is an average size of a cluster of four equivalent water molecules, may no longer be valid in generating the δ_{ij} for -**OH** - W interactions. Therefore, in our optimization for other polar particles, if a similar trend occurs, we optimize δ_{ij} instead of obtaining δ_{ij} according to the LB rules.

Table 7. Experimental and Calculated Hydration Free Energies (kJ/mol) of Organic Compounds Used To Fit Parameters[a]

compound	cal.	exp.
Alkanes		
methane	8.2	8.4[b]
ethane	7.3	7.4
propane	8.2	8.3
isobutane	9.3	9.5
Alcohols		
methanol	−20.1	−20.2[c]
ethanol	−20.5	−21.0
1-propanol	−20.1	−20.4
Ethers		
dimethyl ether	−7.7	−8.0
diethyl ether	−6.2	−6.8
Amines		
methyl amine	−18.3	−19.1
ethyl amine	−19.4	−18.8
dimethyl amine	−19.3	−18.0
diethyl amine	−17.4	−17.0
trimethyl amine	−14.8	−13.6
triethyl amine	−11.4	−12.6
Ketones and Aldehydes		
acetone	−14.1	−16.1
butanone	−15.2	−15.2
Amides		
N-methylacetamide	−41.0	−42.1
acetamide	−39.7	−40.6
N,N-dimethylacetamide	−34.9	−35.6
N-propylguanidine	−44.8	−44.8
Carboxylic Acids		
acetic acid	−27.4	−28.0
propionic acid	−27.8	−27.1
Aromatics		
benzene	0.0	−3.6
toluene	−3.7	−3.7
p-cresol	−29.0	−25.6
pyridine	−20.1	−19.7
Alkyl Ammoniums[d]		
methylammonium	−303	−298
ethylammonium	−285	−286
Carboxylate Ions[d]		
acetate	−338	−334
propionate	−325	−328
Thiols and Sulfides		
methanethiol	−5.7	−5.2
ethanethiol	−4.8	−5.4
methyl methyl sulfide	−8.5	−6.4
methyl ethyl sulfide	−5.6	−6.2

[a] All our calculations were carried out at 300 K. [b] Reference 65. [c] Reference 69. [d] Reference 70.

We also noticed that the inductive effect of heteroatoms could polarize alkyl chains, especially the carbon center that connects to heteroatoms, which could have more favorable interactions with water than with nonpolarized carbons. Thus, we introduced an extra type of carbon (-**CH_x**-P with P for polar group) that replaced the alkyl particles connecting to the -**OH** group. The ϵ_{ij} for W- -**CH_x** was optimized to be 1.15 kJ/mol. In addition to the hydroxyl group, -**CH_x** was also used to replace the alkyl particles that directly connect

Table 8. Optimized Nonbonded Parameters (ϵ_{ij} and δ_{ij}) for Interactions between Protein Particles and CG Water

particle types interacting with CG water	ϵ_{ij} (kJ/mol)	δ_{ij} (nm)
W[a]	5.00	0.47
CH_4	1.60	0.430
-CH_3	1.30	0.430
>CH_2	0.86	0.430
>CH-	0.45	0.430
-OH	14.00	0.280
-O-	14.20	0.280
-NH_2	13.50	0.280
>NH	20.50	0.280
>N-	24.00	0.280
>C=O	0.80	0.415
>C=**O**	5.25	0.375
-CO-**NH**-	5.10	0.400
-CO-**NH_2**	4.50	0.400
-CO-**N**<	4.50	0.400
-CO-**OH**	5.50	0.280
=CH-	0.78	0.4225
=N-	3.30	0.400
=NH	1.00	0.400
-**C**OO^-	2.00	0.415
-**CO**O^-	25.20	0.375
-**NH_3^+**	36.00	0.400
-SH	3.40	0.405
-S-	3.90	0.405

[a] The ϵ_{ii} and δ_{ii} of W−W interaction from the original model by Marrink et al.[18]

to ether, ketone, aldehyde, carboxylic acid, amine, amide, ammonium, and carboxylate ion groups.

The parameters for ether oxygen (-**O**-) were optimized in a similar way as those of -**OH**. The ϵ_{ij} for W- -**O**- was optimized by dimethyl ether and diethyl ether. Just like alcohols, δ_{ij} of the W- -**O**- is reduced to 0.28 nm.

Amines. There are three types of CG particles for amines. -**NH_2**, >**NH** and >**N**- are used to represent primary, secondary, and tertiary amines with its hydrogen, respectively. For each type, the simplest compounds (mono-, di-, or tri-) methyl and ethyl amine were used to optimize the parameters, respectively. Like δ_{ij} of W- -**OH** of alcohol, δ_{ij} of W- -**NH_2**, W- >**NH**, and W- >**N**- are all set to 0.28 nm to reduce the perturbation from neighboring carbon groups.

Ketones, Aldehydes, Carboxylic Acids, and Amides. Acetone and butanone were used to optimize the interaction parameters of carbonyl carbon (>**C**=O) and carbonyl oxygen (>C=**O**) with CG water. For simplicity, the aldehyde functional group (-**CH**=**O**) was considered to be composed of >**C**=O and >C=**O** particles. The parameters for carbonyl carbon and carbonyl oxygen were also applied to carboxylic acid and amide. Thus for amides, the parameters of interaction between amide nitrogen and CG water were next to be optimized. There are three kinds of amide nitrogens in our model: the N-monosubstituted (-CO-**NH**-) group that can be used for backbone amide, the -CO-**NH_2** group that can be used for side chains of Asn and Gln, and the N-disubstituted (-CO-**N**<) group. With the optimized carbonyl carbon and oxygen, we optimized parameters of the -CO-**OH** group of carboxylic acid by acetic acid and propionic acid.

Aromatics. There are two types of aromatic carbon groups. One carries a hydrogen atom and the other carries no hydrogen. We used =**CH**- to represent both of them. The HFEs of benzene, toluene, and p-cresol were used to optimize the parameters for the W- =**CH**- interaction. There are also two types of nitrogen groups on heterocyclic aromatic rings (=**N**- and >**NH**). The one (>**NH**) with hydrogen is a HB donor and the other one (=**N**-) is a HB acceptor. They cannot be represented by just one particle type. Therefore, the parameter for W- =**N**- interaction was obtained based on pyridine. The parameters for heterocyclic >**NH** are presumed to be the same as monosubstituted amide >**NH**. The average absolute deviation of the calculated solvation free energies of all aromatic compounds is 2.59 kJ/mol.

Guanidine. As there are no experimental solvation data for a series of guanidine compounds, we could not systematically optimize the parameters of different nitrogens for Arg side-chain analogue, N-propylguanidine. A guanidine group has three sp^2 nitrogens that connect to a central sp^2 carbon. Two of them have single bonds with the central carbon and the other one has a double bond. We assumed that the nitrogens that have single bonds have the same parameters (ϵ_{ij}, δ_{ij}) as amides (-**NH**-, -**NH_2**). In addition, the parameters of carbonyl carbon (>**C**=O) are used for central carbon in the guanidine group. Then, we optimized the interaction parameters for the nitrogen (=**NH**) that has a double bond.

All the optimized ϵ_{ij} and δ_{ij} are given in Table 8. The optimized ϵ_{ij} of interactions between all heteroatom groups and CG water were much larger than that for carbons. This could be understood since the ϵ_{ij} of these polar groups accounted not only for the vdW interactions with water but also the strong electrostatic interactions with water.

3.5. Calculated Hydration Free Energies of Small Molecules. These parameters allowed us to reproduce experimental HFEs of 105 compounds with an average absolute error of about 1.4 kJ/mol (Tables 7 and 9). In particular, we used 35 compounds as a fitting set to obtain all the parameters (Table 7). The average absolute error for the fitting set is 1.1 kJ/mol. For another 70 compounds as a test set (Table 9), the average absolute error is 1.6 kJ/mol. In addition, the correlation plot between the calculated and experimental HFEs is given in Figure 1. The square of the correlation coefficient (R^2) is 0.9993. Our parameters can reproduce HFEs with reasonable accuracy.

3.6. Comparison of Transfer Free Energies of Side-Chain Analogues of Amino Acids from Cyclohexane to Water. Several calculations use the transfer free energy of side-chain analogues of amino acids from cyclohexane to water to benchmark the accuracy of all-atom force fields.[48−50] With the optimized bonded and nonbonded parameters in hand, we were also able to compute the SFEs of the analogues in these two media and compare the results from experiments and the all-atom simulations (Table 10).

In our calculations, CG water was used to calculate SFEs in water and six >**CH_2** particles were used to build cyclohexane. Here, we calculated 19 amino acid side-chain analogues except for Gly and compared the results with the works by Shirts et al.,[49] Chang et al.,[50] MacCallum and Tieleman,[48] where 14−18 amino acid side-chain analogues were calculated. The analogue for Gly was dihydrogen,[39,40] which was out of the scope of our current CG model. The calculated SFEs of the analogues in water

Table 9. Experimental and Calculated Hydration Free Energies (kJ/mol) of Organic Compounds Used To Test Parameters

compound	cal.	exp.	compound	cal.	exp.
Alkanes			Aromatics		
n-butane	9.2	9.0	o-xylene	−4.3	−3.8
n-pentane	10.2	9.8	naphthalene	−7.2	−10.0
isopentane	9.8	10.0	2-methylpyridine	−22.0	−19.4
n-hexane	10.9	10.7	3-methylpyridine	−22.9	−20.0
isohexane	11.6	10.6	methylindole	−31.0	−24.3
3-methylpentane	11.9	10.5	methylimidazole	−39.9	−42.1
cyclohexane	6.4	5.2	Ethers		
n-heptane	11.4	11.0	methyl isopropyl ether	−8.5	−8.4
methylcyclohexane	10.9	7.2	methyl n-propyl ether	−7.4	−7.0
n-octane	13.3	12.1	ethyl n-propyl ether	−6.0	−7.6
Alcohols			di-n-propyl ether	−3.2	−4.8
2-propanol	−22.0	−19.9	methyl t-butyl ether	−7.8	−9.3
1-butanol	−18.9	−19.8	di-isopropyl ether	−8.0	−2.2
2-butanol	−19.4	−19.1	di-n-butyl ether	−1.9	−3.5
2-methyl-1-propanol	−17.7	−18.9	tetrahydropyran	−10.7	−13.1
t-butanol	−23.9	−18.9	Ketones and Aldehydes		
1-pentanol	−16.7	−18.7	pentan-2-one	−13.9	−14.8
2-pentanol	−17.7	−18.4	pentan-3-one	−14.9	−14.3
3-pentanol	−17.0	−18.2	3-methyl-2-butanone	−15.9	−13.6
3-methyl-1-butanol	−17.5	−18.5	2-hexanone	−12.2	−13.8
2-methyl-2-butanol	−19.6	−18.5	4-methyl-2-pentanone	−11.9	−12.8
1-hexanol	−16.6	−18.2	2-heptanone	−11.1	−12.7
3-hexanol	−15.5	−17.1	4-heptanone	−11.9	−12.3
4-methyl-2-pentanol	−17.4	−15.7	2,4-dimethyl-3-pentanone	−15.7	−11.5
2-methyl-3-pentanol	−15.1	−16.3	acetaldehyde	−15.8	−14.6
2-methyl-2-pentanol	−17.7	−16.4	propanal	−16.9	−14.4
cyclohexanol	−19.9	−22.9	butanal	−15.5	−13.3
2,3-dimethyl-2-butanol	−16.3	−16.4	pentanal	−14.1	−12.7
Amines			hexanal	−12.8	−11.8
propyl amine	−18.0	−18.4	Thiols and Sulfides		
butyl amine	−16.7	−18.0	propanethiol	−4.0	−4.6
pentyl amine	−16.0	−17.2	ethyl ethyl sulfide	−3.9	−6.0
hexyl amine	−15.4	−16.9	Carboxylic Acids		
dipropyl amine	−14.8	−15.3	butyric acid	−26.6	−26.6
dibutyl amine	−11.0	−13.9	Alkyl Ammoniums		
piperidine	−21.8	−21.4	propylammonium	−277	−279
N-methylpiperidine	−14.9	−16.3	butylammonium	−274	−277
Amides			Carboxylate Ion		
propionamide	−39.2	−39.3	butyrate	−321	−324
N-acetylpyrrolidine	−37.1	−40.9			

from the three all-atom simulation studies deviated on average from experimental values by about 3.4, 2.9 and 4.5 kJ/mol, respectively. Taking advantage of the simulation speed of our CG model, we were able to directly optimize our parameters by fitting the SFEs in water. As a result, the average deviation of our simulations was only about 1.3 kJ/mol.

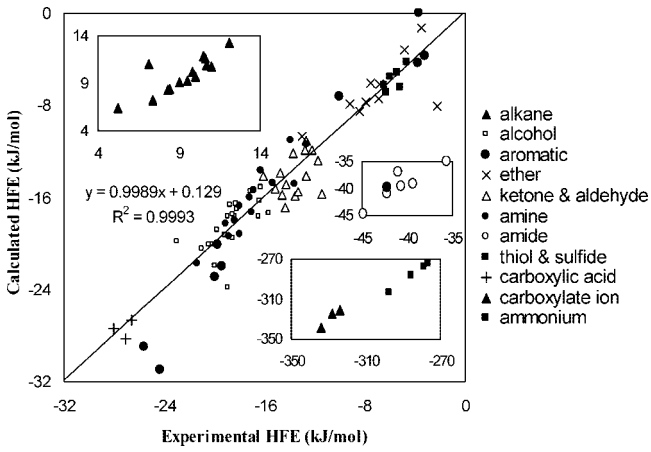

Figure 1. Correlation plot between the calculated and experimental HFEs.

For the SFEs of these amino acid side-chain analogues in cyclohexane, our simulations had an average deviation of about 1.8 kJ/mol from experimental data. This is comparable to the all-atom simulations reported by Chang et al. (2.3 kJ/mol) and by MacCallum et al. (2.4 kJ/mol). Since the SFEs were all determined by interactions between protein particles, our result indicated that most of the assumptions made in the optimization of protein−protein particle interactions might work. However, there were a few exceptions. First, our model systematically underestimated the SFE of hydroxyl-containing molecules (Ser and Thr) by about 3 kJ/mol. Interestingly, even the all-atom model has the same problem, which was pointed out by Chang et al.[50] Second, the ϵ_{ij} of $>\mathbf{CH_2}$ - $=\mathbf{CH}$- interaction from the combination rule was so strong that the SFE of toluene could be exaggerated by more than 15 kJ/mol. The reason may be that our parameters for aromatic carbon were optimized from the self-SFE of pure benzene. In liquid benzene, the

Table 10. Solvation Free Energies of Side-Chain Analogues of Amino Acids in Water or Cyclohexane from Our Calculation, the Simulations by Shirts et al.,[49] by Chang et al.,[50] and by MacCallum and Tieleman[48] Are Compared with Experimental Data[39,40]

	SFE of side-chain analogues in water (kJ/mol)				SFE of side-chain analogues in cyclohexane (kJ/mol)				
	this work	Shirts	Chang	MacCallum	exp.	this work	Chang	MacCallum	exp.
Ala (methane)	8.4	9.0	9.2	9.8	8.4	0.6	1.0	0.9	0.6
Val (propane)	8.2	10.7	10.4	12.0	8.3	−8.7	−6.6	−6.5	−8.6
Leu (isobutane)	9.3	11.3	11.1	13.7	9.5	−14.0	−9.2	−10.3	−11.0
Ile (butane)	9.2	11.2	10.5	12.2	9.0	−11.8	−9.9	−8.8	−11.6
Ser (methanol)	−20.7	−19.7	−19.1	−20.2	−21.2	−3.4	−3.2	−4.2	−6.9
Thr (propanol)	−20.1	−18.7	−19.4	−20.3	−20.4	−6.7	−6.6	−7.2	−9.7
Cys (methanethiol)	−6.4	−1.8	−1.3	−0.5	−5.2	−9.9	−8.9	−9.2	−10.5
Met (methylethylsulfide)	−6.9	−0.3	0.1	−7.1	−6.2	−16.7	−14.5	−14.2	−16.0
Asp (acetic acid)	−29.5			−30.5	−28.0	−11.7		−13.1	−9.2
Glu (propionic acid)	−28.3			−19.0	−27.1	−13.5		−16.6	−13.9
Lys (butylamine)	−16.7			−13.6	−17.9	−15.0		−13.9	−16.4
Arg (N-propylguanidine)	−44.0			−43.9	−44.8	−26.2		−22.7	−20.6
Asn (acetamide)	−40.4	−35.6	−35.7	−34.5	−40.5	−12.9	−13.7	−12.9	−14.3
Gln (propionamide)	−39.2	−35.5	−36.6	−31.4	−39.2	−15.0	−16.8	−17.7	−16.1
Pro (N-acetylpyrrolidine)	−37.1				−40.9	−27.2			−28.2
Phe (toluene)	−3.7	−2.7	−2.9	−1.2	−3.2	−23.0	−19.7	−20.2	−17.5
Tyr (p-cresol)	−29.0	−21.2	−20.5	−18.8	−25.5	−24.9	−22.7	−23.0	−25.0
His (methylimidazole)	−39.9	−36.2	−35.6	−28.0	−42.9	−21.7	−19.0	−19.7	−23.4
Trp (methylindole)	−31.0	−17.3	−24.0	−16.2	−24.6	−34.1	−38.1	−30.6	−34.3
average error (kJ/mol)	1.3	3.4	2.9	4.5		1.8	2.3	2.4	

Table 11. Transfer Free Energies of Side-Chain Analogues of Amino Acids from Cyclohexane to Water from Our Calculations, the Simulations by Chang et al.[50] and by MacCallum and Tieleman,[48] along with Those Determined by Experiments[39,40]

	this work	Chang	MacCallum	exp.
Ala (methane)	7.8	8.2	8.9	7.8
Val (propane)	16.9	17.0	18.5	16.9
Leu (isobutane)	23.3	20.3	24.0	20.5
Ile (butane)	21.0	20.4	21.0	20.6
Ser (methanol)	−17.3	−15.9	−16.0	−14.3
Thr (propanol)	−13.4	−12.8	−13.1	−10.7
Cys (methanethiol)	3.5	7.6	8.7	5.3
Met (methylethylsulfide)	9.8	14.6	7.1	9.8
Asp (acetic acid)	−17.8		−17.4	−18.8
Glu (propionic acid)	−14.8		−2.4	−13.2
Lys (butylamine)	−1.7		0.3	−1.5
Arg (N-propylguanidine)	−17.8		−21.2	−24.2
Asn (acetamide)	−27.5	−22.0	−21.6	−26.2
Gln (propionamide)	−24.2	−19.8	−13.7	−23.1
Pro (N-acetylpyrrolidine)	−9.9			−12.7
Phe (toluene)	19.3	16.8	19.0	14.3
Tyr (p-cresol)	−4.1	2.2	4.2	−0.5
His (methylimidazole)	−18.2	−16.6	−8.3	−19.5
Trp (methylindole)	3.1	14.1	14.4	9.7
average error (kJ/mol)	2.2	2.4	4.2	

aromatic−aromatic interactions and/or the electrostatic interactions are always present between benzene molecules, and they therefore contribute to the optimized values for the **=CH-** - **=CH-** parameters. But no such interactions exist between **>CH$_2$** and **=CH-** particles. The same exaggeration, though not very large, could also be seen in the all-atom models.[50] Thus, the ϵ_{ij} for **>CH$_2$** - **=CH-** interaction was separately optimized. It is 0.34 kJ/mol.

Finally, the transfer free energies of the side-chain analogues from cyclohexane to water were calculated (Table 11). The transfer free energies from our model agree reasonably well with experimental data with an average deviation of 2.2 kJ/mol. These results are comparable to those obtained by all-atom simulations reported by Chang et al. and MacCallum et al., which have average absolute errors of 2.4 and 4.2 kJ/mol, respectively.

3.7. Limitation. We treat the electrostatic interaction in a similar way to that of the CG model of Marrink et al. by implicitly incorporating electrostatic terms into vdW terms. Thus, the long-range electrostatic interaction beyond the cutoff (1.2 nm) is missing. As pointed out before,[18,20] the results for a system where the long-range electrostatics are important may be affected. For a system such as a huge protein that has electrostatic complex structures in its low-dielectrics interior, the results with our model should be interpreted with caution.

4. Conclusion

Further development and parametrization of our coarse-grained protein model in tandem with a CG solvent model have been carried out. Totally 24 types of protein particles were built to describe various amino acids. The interaction parameters between protein particles and between protein and CG water particles were optimized based on the self-solvation free energies of eight molecules and the hydration free energies of a set of 35 molecules. Our model can reproduce hydration free energies of 105 organic compounds with an average deviation of about 1.4 kJ/mol from experimental data. It can also give transfer free energies of the side-chain analogues of all 19 amino acids from cyclohexane to water with an average absolute error of about 2.2 kJ/mol. The results showed that parametrization based on solvation properties enables our model to capture solvation changes in amino acids, which is essential for protein folding. It is possible to incorporate our protein model systematically into a CG solvent model for effective studies of protein folding and protein−protein interactions. Full parametrization in this direction is in progress.

Acknowledgment. We are grateful to RGCHK (N-HKUST 623/04, CA06/07.SC05) and NSFC (20225312) for financial support of the research.

Supporting Information Available: Schematic views of optimized geometries of small molecules by quantum mechanics calculation (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Urbanc, B.; Borreguero, J. M.; Cruz, L.; Stanley, H. E. *Methods Enzymol.* **2006**, *412*, 314–338.

(2) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147.

(3) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860.

(4) Smith, A. V.; Hall, C. K. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 344.

(5) Marchut, A. J.; Hall, C. K. *Proteins* **2007**, *66*, 96–109.

(6) Akkhipov, A.; Freddolino, P. L.; Schulten, K. *Structure* **2006**, *14*, 1767.

(7) Gô, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183.

(8) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844–11849.

(9) Thirumalai, D.; Guo, Z. *Biopolymers* **1995**, *35*, 137.

(10) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616–11629.

(11) Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394–408.

(12) Sharma, S.; Ding, F.; Dokholyan, N. V. *Biophys. J.* **2007**, *92*, 1457–1470.

(13) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(14) Izvekov, S.; Voth, G. A. *J. Chem. Theor. Comput.* **2006**, *2*, 637.

(15) Ayton, G. S.; Voth, G. A. *J. Struct. Biol.* **2007**, *157*, 570.

(16) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045.

(17) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.

(18) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.

(19) Monticelli, L.; Kandasamy, S.; Periole, X.; Larson, R.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819.

(20) Marrink, S. J.; Risselada, H, J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys Chem. B* **2007**, *111*, 7812.

(21) Baron, R.; Trzesniak, D.; de Vries, A. H.; Elsener, A.; Marrink, S. J.; van Gunsteren, W. F. *Chemphyschem* **2007**, *8*, 452.

(22) Kasson, M. P.; Kelly, N. W.; Singhal, N.; Vrljic, M.; Brunger, A. T.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11916.

(23) Marrink, S. J.; Mark, A. E. *J. Am. Chem. Soc.* **2003**, *125*, 15233.

(24) de Vries, A. H.; Mark, A. E.; Marirnk, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 4488–4489.

(25) de Vries, A. H.; Yefimov, S.; Mark, A. E.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5392.

(26) Bond, P. J.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2006**, *128*, 2697.

(27) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. *J. Phys. Chem. B* **2006**, *110*, 3674–3684.

(28) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J. Am. Chem. Soc.* **2007**, *129*, 10126.

(29) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. *J. Phys. Chem. B* **2007**, *111*, 11095.

(30) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. *J. Struct. Biol.* **2007**, *157*, 593.

(31) Scott, K. A.; Bond, P. J.; Ivetac, A.; Chetwynd, A. P.; Khalid, S.; Sansom, M. S. P. *Structure* **2008**, *16*, 621.

(32) Sansom, M. S. P.; Scott, K. A.; Bond, P. J. *Biochem. Soc. Trans.* **2008**, *36*, 27.

(33) Honig, B.; Yang, A. S. *Adv. Protein Chem.* **1995**, *46*, 27–58.

(34) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.

(35) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

(36) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.

(37) Fauchere, J.-L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369.

(38) Kim, A.; Szoka, F. C. *Pharm. Res.* **1992**, *9*, 504.

(39) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664.

(40) Wolfenden, R.; Anderson, L.; Cullis, P.; Southgate, C. *Biochemistry* **1981**, *20*, 849.

(41) Wimley, W. C.; Creamer, T. P.; White, S. H. *Biochemistry* **1996**, *35*, 5109.

(42) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.

(43) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276.

(44) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569–1489.

(45) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(46) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(47) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656.

(48) MacCallum, J. L.; Tieleman, D. T. *J. Comput. Chem.* **2003**, *24*, 1930.

(49) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(50) Chang, J.; Lenhoff, A. M.; Sandler, S. I. *J. Phys. Chem. B* **2007**, *111*, 2098.

(51) Han, W.; Wu, Y.-D. *J. Comput. Theory Chem.* **2007**, *3*, 2146.

(52) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730–748.

(53) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.

(54) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.

(55) Berendsen, H. J. C.; Pstma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(56) Mezei, M.; Beveridge, D. L. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1.

(57) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Greber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529.

(58) Frisch, M. J. et al. *Gaussian 03, revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.

(59) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 4244.

(60) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91.

(61) Kitagawa, T.; Miyazawa, T. *Bull. Chem. Soc. Jpn.* **1968**, *41*, 1976.

(62) Herrebout, W. A.; van der Veken, B. J.; Wang, A.; Durig, J. R. *J. Phys. Chem.* **1995**, *99*, 578.

(63) Henderson, D.; Leonard, P. J. Liquid Mixture. In *Physical Chemistry: An Advanced Treatise*; Academic Press: London, 1971; Vol *VIIIB*.

(64) Lide, D. R. In *CRC Handbook of Chemistry and Physics*, 2nd ed.; CRC Press: Boca Raton, FL, 1992.

(65) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.

(66) Schurman, I.; Boord, C. E. *J. Am. Chem. Soc.* **1933**, *55*, 4930.

(67) Osborne, D. W.; Doescher, R. N.; Yost, D. M. *J. Am. Chem. Soc.* **1942**, *64*, 169.

(68) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. *J. Mol. Biol.* **1999**, *290*, 253.

(69) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.

(70) Kang, Y. K.; Nemethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4118.

CT800184C