JCTC Journal of Chemical Theory and Computation

Polarizable Force Field for Protein with Charge Response Kernel

Miho Isegawa and Shigeki Kato*

Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

Received June 9, 2009

Abstract: We present a molecular mechanical force field for polypeptides and proteins involving the electronic polarization effect described with the charge response kernel. All of the electrostatic parameters for 20 amino acids are obtained by ab initio electronic structure calculations and combined with the AMBER99 force field. The refittings of dihedral angle parameters in the torsional potentials are performed so as to reproduce the ab initio optimized geometries and relative energies for the conformers of dipeptides. The present force field is applied to molecular dynamics simulation calculations of the extended alanine tetra and cyclic pentapeptides in aqueous solution. The infrared spectra are calculated in order to analyze the charge polarization effect on the spectral profiles.

1. Introduction

Molecular dynamics (MD) simulation is a powerful means for studying statistical and dynamic properties of biological systems. In these studies, the reliability of calculated results is governed by the quality of employed molecular mechanical (MM) force fields. The standard MM force fields such as AMBER,[1−3] CHARMM,[4−6] GROMOS,[7−9] and OPLS[10−12] have been developed and widely used for MD simulations of proteins.

In order to construct more realistic force fields, the charge polarizable models have been proposed in recent years[13−22] and applied to various biological systems.[23−27] The charge polarization induced by intra- or intermolecular interaction has been represented by the site point dipole,[13−15] fluctuating charge,[16−18] and Drude oscillator[19−21] methods. Thus, the AMBER, CHARMM, and OPLS force fields have been refined with combining the polarizable electrostatic interaction.[14,15,17,21]

In the present work, we develop a charge-polarizable MM model for peptides and proteins based on the charge response kernel (CRK) method.[28−31] The advantage of the CRK model is to be able to describe the nonlocal charge polarization through the off-diagonal elements, which is important for representing the charge migration through peptide bonds and aromatic residues. It is also noted that the CRK matrix elements can be directly determined by ab inito electronic structure calculations for isolated molecules. We employ the second-order Møller−Plesset perturbation (MP2) method for calculating the CRK matrices in this study.[31]

The CRK model has been applied to water, aprotic solvents, and organic radicals, and the infrared (IR) and Raman spectra as well as the diffusion constants in the liquid phase have been calculated.[31−33] Although the applications of the CRK model have been limited to small molecules so far, we extend it for use with polypeptides and proteins in the present study. For this purpose, we devise a method to construct the CRK matrix of protein using the matrices of fragment amino acids.

We combine the present CRK model with the AMBER99[3] potential functions for bond stretching, bending, and torsional and nonbonding van der Waals interactions. As is known, the torsional potentials are strongly correlated to the intramolecular electrostatic interaction and play an important role in determining the conformational stability.[34,35] We therefore readjust the parameters of torsional potentials so as to reproduce the ab initio geometries and relative energies for various conformers of dipeptides. Such a procedure was already employed by Kaminski et al.[34] for improving the OPLS-AA force field, which shows good performance for 20 amino acid dipeptides, and by Wang et al.[35] in construct-

* Corresponding author e-mail: shigeki@kuchem.kyoto-u.ac.jp.

10.1021/ct900295u CCC: $40.75
Published on Web 09/24/2009

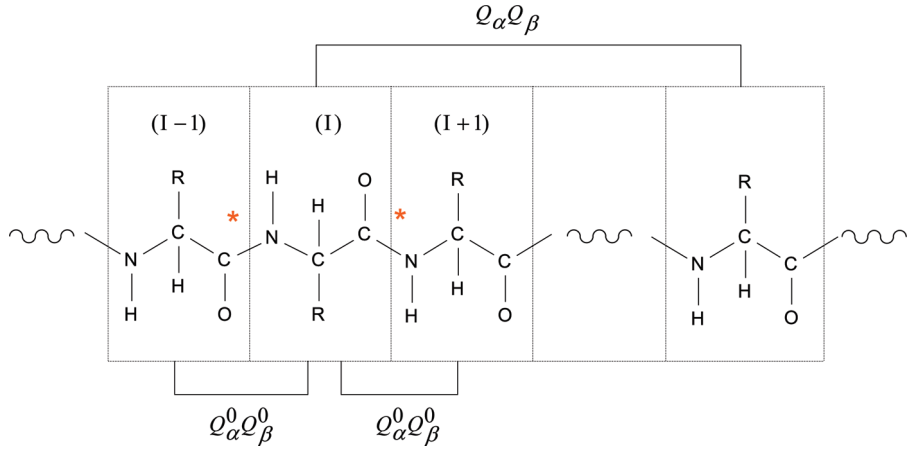

Figure 1. Illustration of electrostatic interactions. The interactions are classified into two patterns: interactions by the reference charges in the nearest neighbor units and (ii) interactions, except for case i, by the polarized charges.

ing a polarizable AMBER force field for alanine dipeptide based on the site point dipole model.

To assess the utility of the present CRK-based MM model, we carry out MD calculations for the extended alanine tetrapeptide and cyclic pentapeptide in aqueous solution. Alanine is the most representative amino acid because of its simplicity and common characteristics with 20 amino acids. Actually, alanine polypeptides have been the subject of many theoretical studies, including the electronic structure and MD simulation methods.[36−41] We calculate the IR spectra of the extended and cyclic peptides because the charge distribution induced by the intra- and intermolecular electrostatic potential is expected to be different between the two forms. The IR spectra for amide modes have been investigated, and the effect of hydrogen bonding on the spectral profiles has been examined.[42−48] Bouř and Keidering[47] carried out MD simulations to calculate the IR intensity of the amide I band for linear alanine pentapeptide in aqueous solution. More recently, Torii[48] calculated the IR and Raman spectra of tetrapeptide in aqueous solution and analyzed the profile of the amide I band in detail. Here, we examine the polarization effects on the calculated spectra by decomposing the spectral shape into the contributions from the intra- and intermolecular electrostatic interaction.

In the next section, we construct the CRK-based polarizable MM model for peptides and proteins combined with the AMBER force field. The electrostatic parameters are determined for 20 amino acids and terminal species, referring to the results of ab initio electronic structure calculations. We optimize the Fourier coefficients in dihedral angle terms so as to reproduce the ab initio relative energies and geometries for the conformers of dipeptides. The results are presented for alanine di- and tetrapeptide in section 2, and for the other dipeptides in the Supporting Information. In section 3, we apply the present force field to MD simulations of the extended alanine tetrapeptide and cyclic pentapeptide in aqueous solution. The effect of intra- and intermolecular hydrogen bonding on the spectral profiles is discussed. The concluding remarks of the present work are summarized in section 4.

2. Force Field

2.1. Electrostatic Potential Function. In the present polarizable force field for polypeptide with the CRK model, the partial charge $Q_{\alpha I}$ at the atomic site α in the *I*th amino acid unit is defined by

$$Q_{\alpha I} = Q^0_{\alpha I} + \sum_{\beta \in J}^{N_J} \sum_{J}^{N_u} K_{\alpha I,\beta J} V^{\mathrm{eff}}_{\beta J} \qquad (1)$$

where N_u is the number of the amino acid units in the polypeptide and N_J is the number of the atomic sites in the unit *J*. $Q^0_{\alpha I}$ is the reference charge initially assigned on the atomic site (see Figure 1). The induced charge is given by the second term of eq 1, where $K_{\alpha I,\beta J}$ is the CRK matrix element and $V^{\mathrm{eff}}_{\beta J}$ is the effective electrostatic potential acting on the β site in the *J*th unit. Note that the CRK matrix is related to the polarizability as

$$\alpha_{mn} = -\sum_{\alpha \in I}^{N_I} \sum_{I}^{N_u} \sum_{\beta \in J}^{N_J} \sum_{J}^{N_u} K_{\alpha I,\beta J} R_{m,\alpha I} R_{n,\beta J} \qquad (2)$$

where $R_{m,\alpha I}$ is the Cartesian coordinate of the atomic site α in the unit *I* with $m = x, y, z$.

In the conventional CRK approach, the reference charge $Q^0_{\alpha I}$ is defined to be the atomic charge of isolated molecule, and only the intermolecular contribution to the electrostatic potential is taken into account for calculating $V^{\mathrm{eff}}_{\beta J}$. However, it is impractical to employ such a method to represent the atomic site charges of the whole polypeptide because the charges are affected by the intramolecular hydrogen bonding and thus strongly depend on the conformation of the polypeptide. In order to avoid such a difficulty, we obtained the atomic charges for each amino acid unit in the isolated condition and assigned them to the reference charges. The charge redistribution due to the interaction between different amino acid units is thus given by the second term of eq 1.

The CRK matrix of the whole polypeptide was also constructed from the matrices obtained for the constituent amino acids. The CRK matrix for glycine tripeptide was used to represent the matrix elements corresponding to the amide bond. Since the CRK matrix must have one zero eigenvalue

with the eigenvector corresponding to the uniform potential, we applied the projection operator $\mathbf{P} = \mathbf{I} - 1/N_a\mathbf{1}$, with N_a being the number of total atomic sites in the polypeptide, as

$$\mathbf{K} = \mathbf{P}\tilde{\mathbf{K}}\mathbf{P}^t \tag{3}$$

where $\tilde{\mathbf{K}}$ is the CRK matrix constructed from those of the fragments. Here, $\mathbf{I}$ is the unit matrix and $\mathbf{1}$ is the matrix with all elements of 1. With this operation, the total charge of the polypeptide is conserved.

The effective electrostatic potential acting on the β site in the Jth unit has both the intra- and intermolecular contributions as mentioned above. The intramolecular part is given by

$$V_{\beta J}^{\mathrm{eff}} = \sum_{\gamma \in L}^{N_L} \sum_{L}'' Q_{\gamma L} \frac{f(|\mathbf{R}_{\beta J} - \mathbf{R}_{\gamma L}|)}{|\mathbf{R}_{\beta J} - \mathbf{R}_{\gamma L}|} \tag{4}$$

where the index L runs over all of the amino acid units except for J and $J \pm 1$, and $f(|\mathbf{R}_{\beta J} - \mathbf{R}_{\gamma L}|)$ is the damping function introduced to attenuate the electrostatic potential at the short range. We omitted the electrostatic contributions from the sites in the nearest neighbor units because the sites in more distant units become important in characterizing the conformation of polypeptide. The site charges are determined so as to satisfy eqs 1 and 4 simultaneously.

Using the intramolecular electrostatic potential defined by eq 4, the electrostatic interaction part of the potential function for an isolated polypeptide is written as

$$\begin{aligned} U_{\mathrm{ES}} = & \frac{1}{2}\sum_{\alpha \in I}^{N_I}\sum_{I}^{N_u}\sum_{\beta \in J}^{N_J}\sum_{J}' Q_{\alpha I}^0 Q_{\beta J}^0 \frac{f_{\alpha I,\beta J}^0}{|\mathbf{R}_{\alpha I} - \mathbf{R}_{\beta J}|} \\ & + \frac{1}{2}\sum_{\alpha \in I}^{N_I}\sum_{I}^{N_u}\sum_{\beta \in J}^{N_J}\sum_{J}'' Q_{\alpha I} Q_{\beta J} \frac{f(|\mathbf{R}_{\alpha I} - \mathbf{R}_{\beta J}|)}{|\mathbf{R}_{\alpha I} - \mathbf{R}_{\beta J}|} \\ & - \frac{1}{2}\sum_{\alpha \in I}^{N_I}\sum_{I}^{N_u}\sum_{\beta \in J}^{N_J}\sum_{J}^{N_u} K_{\alpha I,\beta J} V_{\alpha I}^{\mathrm{eff}} V_{\beta J}^{\mathrm{eff}} \end{aligned} \tag{5}$$

where the first term represents the interactions between the reference charges on the nearest neighboring units. The factor $f_{\alpha I,\beta J}^0$ is introduced to follow the interaction scheme in AMBER; the 1−2 and 1−3 nonbonding interactions are omitted, and the 1−4 interaction is estimated by multiplying the factor of 1/1.2. The second term includes all of the interactions between the amino acid units except for the nearest neighbors. Note that the charge polarization effect due to the intramolecular electrostatic interaction is incorporated in this term. Finally, the third term corresponds to the electronic reorganization energy originated from the charge polarization in the polypeptide.

We employed Thole's damping function[49] for $f(R)$ in eqs 4 and 5,

$$f(R) = \begin{cases} \left(\frac{R}{S}\right)^4 - 2\left(\frac{R}{S}\right)^3 + 2\left(\frac{R}{S}\right) & (R < S) \\ 1 & (R > S) \end{cases} \tag{6}$$

with

$$S = A(\alpha_\alpha \alpha_\beta)^{1/6} \tag{7}$$

Here, R is the distance between the atomic sites α and β, and the critical distance S is determined by the factor A and the atomic polarizability α_α. We set the factor A to be 2.8,[31,33] which was used for the intermolecular interaction in liquid water and aprotic solvent systems, and the polarizalibities as $\alpha_H = 0.514$, $\alpha_C = 1.405$, $\alpha_{CA} = 1.671$, $\alpha_{CT} = 1.822$, $\alpha_N = 1.105$, $\alpha_{NA} = 1.296$, $\alpha_O = 0.862$, and $\alpha_S = 2.540$ in Å^3.[49,50] Here, the subscripts CA, CT, and NA mean the carbonyl carbon, sp^3 carbon, and amide nitrogen, respectively.

2.2. Electrostatic Parametrization. We carried out ab initio calculations for 20 neutral and five charged amino acids at the MP2 level of theory with the 6-31G** basis set. The C-terminal N-methyl group (NME) and N-terminal acetyl group (ACE) were also considered. In the calculations, both the terminals of amino acid were blocked with methyl groups, as shown in Figure 1 with the asterisk. However, the methyl groups were replaced with the hydrogen atoms for relatively large amino acids (i.e., arginine, lysine, phenylalanine, tyrosine, and tryptophan) in order to reduce the calculation cost.

We first performed the geometry optimizations for all of the species at the MP2 level with the 6-31G** basis set. At these optimized geometries, the electrostatic potential (ESP) charges on the atomic sites were obtained by[31]

$$Q_\alpha = \frac{\partial E^{\mathrm{MP2}}}{\partial V_\alpha} \tag{8}$$

where E^{MP2} is the MP2 energy and V_α is the electrostatic potential acting on the site α. We modified the ESP charges by imposing the following three conditions. First, the site charges such as on the hydrogens of the methyl group and the symmetrical carbons and hydrogens in the aromatic ring were replaced with the averaged values. Second, the charges in the amide part were taken from the averaged values of glycine dipeptide at three stable conformers, C7, C5, and α:[51] −0.473 for N, 0.311 for H, 0.580 for C, and −0.500 for O. Finally, the charges of atoms in the side chain were scaled so that the value of the total charge becomes an integer, that is, 0 for neutral and ±1 for charged amino acids, under the condition that the charges of N, H, C, and O in the amide part are fixed. The parameters for the terminal capping groups, NME and ACE, were also determined by taking the average of the charges at three conformers of glycine dipeptide as in the case of the amide part. The reference charges thus obtained are summarized in Tables S1−S27 in the Supporting Information.

We also executed ab initio calculations to obtain the CRK matrices of all of the species defined by[31]

$$K_{\alpha\beta} = \frac{\partial^2 E^{\mathrm{MP2}}}{\partial V_\alpha \partial V_\beta} \tag{9}$$

The CRK matrix has large negative values at the diagonal elements, and the magnitudes of off-diagonal elements rapidly decrease with increasing the number of bonds separating the two sites. Thus, the 1−4 elements become negligibly smaller compared to the diagonal ones. We therefore truncated the CRK matrices at the 1−3 elements to reduce the conformational dependency. It is however noted

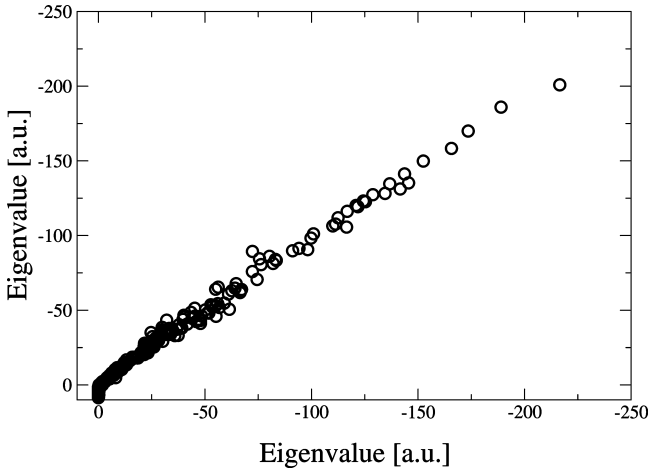

Figure 2. Correlation of eigenvalues between original and truncated CRK matrices for glycine dipeptide and 20 amino acids. Horizontal and vertical axes are for original and truncated matrices.

that the elements beyond the 1−3 of the aromatic ring are not small because of π conjugation and cannot be ignored. For this reason, we retained all of the elements in the aromatic ring for phenylalanine and tyrosine, as well as the elements between the α carbon and aromatic sites for tryptophan. All of the CRK matrices are listed in Tables S1−S27 (Supporting Information). Figure 2 shows the eigenvalues of truncated CRK matrices compared with the original eigenvalues for a glycine dipeptide, and 19 neutral and five charged amino acids, where we can see that the truncated eigenvalues are very close to the original ones.

2.3. Parametrization of Dihedral Potential. The intramolecular potential function for isolated polypeptide is the same as the AMBER force field except for the electrostatic term U_{ES},

$$U = \sum_{\text{bonds}} K_r(R - R_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} V_n[1 + \cos(n\phi - \gamma)] + \sum_{\alpha>\beta}\left[\frac{A_{\alpha\beta}}{R_{\alpha\beta}^{12}} - \frac{B_{\alpha\beta}}{R_{\alpha\beta}^{6}}\right] + U_{ES} \quad (10)$$

Here, the first two terms represent the harmonic bond stretching and bending motions, respectively. The third term is the torsional potential written by the Fourier series. The last two terms are the nonbonding van der Waals and electrostatic interactions.

Since the stretching and bending terms are rather independent of the electrostatic term, we adopted the same parameters as those for AMBER99 for these terms. However, the dihedral terms strongly correlate to the electrostatic term and are responsible to the conformational stability. We therefore reparametrized the torsion parameters V_n so as to reproduce the relative energies and geometries of several stable dipeptide conformers obtained with the ab inito electronic structure calculations. In order to determine the parameters for the backbone dihedral angles (ϕ, ψ), alanine dipeptide was selected because the effect of the side chain on the backbone conformation is small. We generated 1296

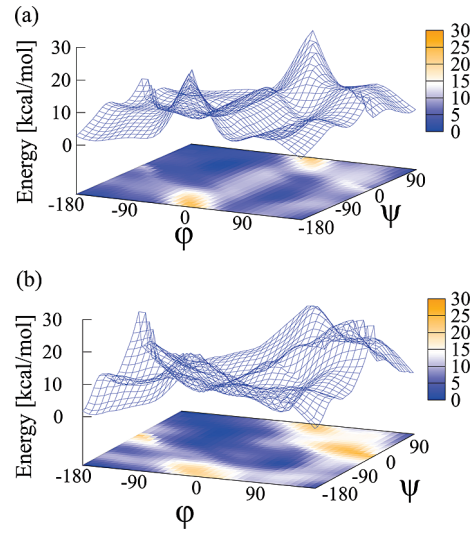

Figure 3. Potential energy surfaces as a function of (ϕ, ψ) obtained by (a) ab initio calculation and (b) the CRK model. The most stable conformer $C7_{eq}$ is taken to be 0 kcal/mol.

reference grid points in total (36 points are taken for each dihedral angle) and performed ab initio calculations at the MP2 level with the 6-31G** basis set. The ab initio calculated quantum mechanical (QM) potential surface is shown in Figure 3a as a function of the dihedral angles (ϕ, ψ). The Fourier coefficients of torsional potential are obtained by minimizing the function:

$$R = \sum_{i}^{\text{ngrid}} \omega_i(\Delta E_i^{MM} - \Delta E_i^{QM})^2 \quad (11)$$

where the index i refers a grid point and ω_i is the weighting factor

$$\omega_i = \exp(-\zeta \Delta E_i^{QM}) \quad (12)$$

with

$$\Delta E_i^{QM} = E_i^{QM} - E_{C7_{eq}}^{QM} \quad (13)$$

and

$$\Delta E_i^{MM} = (E_i^{MM} - E_i^{torsion}) - (E_{C7_{eq}}^{MM} - E_{C7_{eq}}^{torsion}) \quad (14)$$

Here, $E_i^{torsion}$ is the torsion energy corresponding to the backbone dihedral angles ϕ and ψ, and the subscript $C7_{eq}$ is the conformer with the angles ϕ and $\psi = -85.8°$ and 79.2°, which is the most stable conformer of dipeptide with the intramolecular hydrogen bond. The parameter ζ in the weighting factor is determined so as to minimize the relative energy rms deviation from the ab initio reference values for the five stable conformers: $C7_{eq}$, C5, $C7_{ax}$, α_L, and α'.

The optimized torsional parameters of the main chain determined for alanine dipeptide were used for most of the other dipeptides. However, when either the relative energy or the optimized geometry extremely differs from the ab initio reference, the least-squares fittings of dihedral angle parameters for the side chain were performed. In this fitting,

ab initio grid points for the side chain are generated in a crosslike manner for each stable conformer.[34]

2.4. Evaluation of the Force Field for Isolated Peptide. The fitted dihedral parameters are summarized in Table S28 in the Supporting Information. To evaluate the present polarizable MM force field, we compared the relative energies and key dihedral angles at the stable conformations of all of the dipeptides with those obtained by ab initio calculations. For this purpose, we carried out the geometry optimizations of the stable conformers for 20 neutral and five charged dipeptides at the HF level with the 6-31G** basis set, starting from the initial geometries given by Kaminski et al.[34] At the optimized geometries, the MP2 calculations were repeated to estimate the relative energies. The resultant dihedral angles are summarized in Tables S29−S46 (Supporting Information).

We next carried out the geometry optimizations of the conformers for each peptide with the present polarizable MM model, where the ab initio optimized geometries were used as the initial guess. The calculations were also performed with the original AMBER99 force field. Since there are several stable conformers for dipeptides, the rms deviation from the ab initio reference energies was calculated by

$$R_{\mathrm{E}} = \sqrt{\frac{1}{N_{\mathrm{C}}}\sum_{i}^{N_{\mathrm{C}}}(E_i^{\mathrm{MM}} - E_i^{\mathrm{QM}})^2} \qquad (15)$$

where N_{C} is the number of conformers and E_i^{MM} and E_i^{QM} correspond to the MM and ab initio energies, respectively. The rms deviation of the dihedral angles was also given by

$$R_{\mathrm{D}} = \sqrt{\frac{1}{N_{\mathrm{C}}N_{\mathrm{D}}}\sum_{i}^{N_{\mathrm{C}}}\sum_{j}^{N_{\mathrm{D}}}(\Psi_{ij}^{\mathrm{MM}} - \Psi_{ij}^{\mathrm{QM}})^2} \qquad (16)$$

where N_{D} is the number of key dihedral angles for a peptide and Ψ_{ij}^{MM} and Ψ_{ij}^{QM} are the dihedral angles at the MM and ab initio optimized geometries, respectively. It is noted that there were conformers whose stable geometries show very large deviations from the corresponding ab initio references for some peptides. For such a case, we excluded the results of these conformers from the estimations of rms deviations.

Alanine Dipeptide and Tetrapeptide. Figure 3b shows the potential energy surface as a function of ϕ and ψ for alanine dipeptide and its contour map calculated with the present MM model. We used the same geometries as for the calculation of the ab initio surface, Figure 3a. Compared with the ab initio surface, we can see that the low potential energy regions involving the conformers $C7_{eq}$, C5, and $C7_{ax}$ are well reproduced, while the high-energy regions, ($-180° \le \phi \le -140°$, $-30° \le \psi \le 40°$) and ($50° \le \phi \le 180°$, $-10° \le \psi \le 100°$), are too high. This is because we employed the weighting factor, which emphasizes low-energy grid points in determining the torsional potential parameters. Table 1 shows the rms deviations of relative energies and dihedral angles for alanine dipeptide. The zero of ab initio relative energy was that of the most stable conformer $C7_{eq}$, while the MM energies were shifted so as to minimize the rms deviation. The five conformers, $C7_{eq}$, C5, $C7_{ax}$, α_L, and α', are reproduced well, but there is no minimum corresponding to β_2 on the MM potential surface. This is because the β_2 minimum exists at the high-energy regions in Figure 3, and the shallow minimum corresponding to this conformer disappears on the potential surface with the present force field. As seen in Table 1, the original AMBER force field reproduces the four conformers $C7_{eq}$, C5, $C7_{ax}$, and β_2, but the conformers α_L and α' are not reproduced. Wang et al.[35] optimized the atomic charges and dihedral parameters for alanine dipeptide taking into account the polarization effect with the AMBER02 force field. However, the optimized force field failed to reproduce some of the stable conformers. Those results indicate that we may need to use a more elaborate form of the potential function including the coupling terms between the torsional angles in order to reproduce all of the conformers.

Table 1. Alanine Dipeptide Relative Energies [kcal/mol] and RMS Deviations of Backbone Dihedral Angles [deg] from the ab Initio References[a]

conformer	ab initio[b]	(ϕ, ψ)	CRK	AMBER
$C7_{eq}$	0.00	(−85.7, 79.1)	−0.03/12.4	0.25/29.5
C5	1.33	(−157.2, 159.8)	1.31/14.1	1.15/5.9
$C7_{ax}$	2.38	(76.0, −55.4)	2.32/15.4	2.48/20.9
β_2	3.00	(−130.9, 22.3)		2.83/1.7
α_L	4.42	(67.0, 30.2)	4.51/30.8	
α'	5.98	(−166.0, −39.9)	5.99/22.6	
rmsd			0.05/20.2	0.18/18.3

[a] RMS deviations are in the bottom. [b] MP2 6-31G**//HF6-31G**.

Table 2. Alanine Tetrapeptide Relative Energies [kcal/mol] and RMS Deviations of Backbone Dihedral Angles [deg] from the ab Inito References[a]

conformer	ab inito[b]	CRK	AMBER
1	5.64	4.67/13.6	4.92/6.3
2	5.28	4.58/15.2	5.16/18.0
3	0.00	1.04/12.4	2.93/18.6
4	6.77	4.45/15.9	
5	7.18	4.98/14.7	4.41/8.1
6	2.39	4.63/23.4	2.25/22.3
7	6.45	6.21/9.9	
8	4.63	7.01/16.2	7.53/19.6
9	7.91	8.65/17.2	7.71/23.9
10	7.40		5.51/30.1
rmsd		1.63/15.8	1.90/19.8

[a] RMS deviations are in the bottom. [b] MP2 6-31G**//HF6-31G**.

The transferability of the potential parameters determined for dipeptides to larger polypeptides is crucial in constructing a MM force field for proteins. We applied the present parameters to alanine tetrapeptide. Table 2 shows the relative energies at the optimized geometries and the rms deviations of dihedral angles for 10 conformers[52] obtained with the present and AMBER force fields. The numbering of the conformers was taken from ref 52. Conformer **3** is the most stable with the ab initio and present force fields because of the existence of intramolecular hydrogen bonding, while conformer **6** is more stable than conformer **3** with the AMBER force field. Conformer **10** with high energy was not reproduced with the present model. The angles ϕ_1 and ψ_1 at the converged geometry largely deviate from the ab initio values. On the other hand, the AMBER force field failed to reproduce the ab initio geometries of conformers **4** and **7**.

Asparagine and Glutamine Dipeptides. Both of the peptides have an amide group in the side chain which participates in the formation of intramoleular hydrogen bonds. Actually, the most stable ab initio conformer of asparagine is stabilized by the two intramolecular hydrogen bonds between the atoms in the backbone and side chain with bond distances of 2.34 and 2.14 Å, respectively. For the asparagine dipeptide, the ab initio stable conformers are reproduced well by the present model with the relative energy and dihedral rms deviations, 0.43 kcal/mol and 15.8°, respectively. For the glutamic acid, two of the 11 conformers obtained by the ab initio calculations are not reproduced with both the present and AMBER force fields. Note that we used the original AMBER parameters for the dihedral angles for the side chains for both the peptides. The results of the two peptides are shown in Tables S47 and S48 (Supporting Information).

Valine, Isoleucine, and Leucine Dipeptides. The side chains of these peptides are composed of alkyl groups. For the valine and isoleuicine dipeptides, all of the ab initio conformers are well reproduced. The resultant rms deviations of both the peptides are ∼0.4 kcal/mol for the relative energies and ∼9.0° for the dihedral angles, as shown in Tables S49 and S50 (Supporting Information). However, there are large deviations of dihedral angles of the main chain from the ab initio values for conformers **3**, **6**, and **7** of leucine, as shown in Table S51 (Supporting Information).

Serine and Threonine Dipeptides. A hydroxyl group is included in the side chains of both the peptides. For the serine dipeptide, we refitted the dihedral parameters for the side chain, χ_1 (N−CT−CT−OH), χ'_1 (OH−CT−CT−C), and χ_2 (CT−CT−OH−HO), keeping the main chain torsion parameters fixed. The rms deviation of 1.55 kcal/mol for the relative energy and a value of 9.5° for the dihedral angle are comparable with those of the AMBER force field, 1.72 kcal/mol and 8.1°, respectively, as shown in Table S52 (Supporting Information). Note that, for some conformers of threonine, the present ab inito optimized dihedral angles differ from those of the corresponding conformers in ref 34. For both force fields, the optimization of conformer **2** resulted in conformer **3**, whose dihedral angles are close to those of conformer **2** except for χ_2. The resultant rms deviations are smaller than the AMBER for both the relative energies and dihedral angles, as shown in Table S53 (Supporting Information).

Phenylalanine, Tyrosine, and Tryptophan Dipeptides. The side chains of these peptides involve an aromatic ring. The rms deviations of relative energies and dihedral angles are small for the phenylalanine and tyrosine dipeptides, as seen in Tables S54 and S55 (Supporting Information). For the tryptophan dipeptide, we executed the fittings of dihedral angle parameters for the side chain, χ_1 (N−CT−CT−C*) and χ'_1 (C*−CT−CT−C). Although two of the nine conformers with high energies could not be described with the present force field, the present model gives smaller rms deviations than the AMBER force field,as shown in Table S56 (Supporting Information).

Cysteine and Methionine Dipeptides. These three peptides involve a sulfur atom in the side chains. For the cysteine dipeptide, the refitting of dihedral parameters for the side chains, χ_1 (N−CT−CT−SH), χ'_1 (SH−CT−CT−C), and χ_2 (CT−CT−SH−HS), was performed. As shown in Table S57 (Supporting Information), the highest-energy conformer among the five stable conformers is not reproduced, but the remaining four conformers are well reproduced with a rms deviation of 0.60 kcal/mol for the relative energy and 14.6° for the dihedral angle. For the methionine, the dihedral refittings for the side chain were not required because the ordering of the relative energy is consistent with the ab initio one with relative energy and dihedral angle rms deviations of 0.82 kcal/mol and 11.0°, respectively. The results are given in Table S58 (Supporting Information).

Charged Asparatic Acid, Glutamic Acid, Histidine, Arginine, and Lysine Diepetides. The asparatic and glutamic acids are negatively charged, while the histidine, arginine, and lysine are positively charged. For these charged peptides, since the geometries of stable conformers in solution deviate from those in the gas phase, we carried out geometry optimization in solution, as performed by Kaminski et al.[34] and used the gas-phase energies at these geometries for the fitting of parameters of the torsional potentials. For the arginine, the dihedral fitting for the side chain, χ_1 (N−CT−CT−CT) and χ'_1 (CT−CT−CT−C), was carried out, and for the other peptides, the dihedral parameters for the side chain were taken from the AMBER force field. The ordering of relative energies is reproduced well except for histidine, and the rms deviations are comparable with those of the AMBER force field. The results are shown in Tables S59−S63 (Supporting Information).

3. Application to Alanine Polypeptides

3.1. MD Simulation Method. The potential energy function of the solvated polypeptide system is expressed as

$$U^{\mathrm{sol}} = U + U' + \sum_{i>j}^{N_{\mathrm{v}}}\sum_{a,b}^{N_{\mathrm{s}}}\left[u_{\mathrm{LJ}}(|\mathbf{r}_{ai}-\mathbf{r}_{bj}|) + Q_{ai}Q_{bj}\frac{f(|\mathbf{r}_{ai}-\mathbf{r}_{bj}|)}{|\mathbf{r}_{ai}-\mathbf{r}_{bj}|}\right] + \sum_{i}^{N_{\mathrm{v}}}\sum_{a}^{N_{\mathrm{s}}}\sum_{\alpha\in I}^{N_I}\sum_{I}^{N_{\mathrm{u}}}\left[u_{\mathrm{LJ}}(|\mathbf{r}_{ai}-\mathbf{R}_{\alpha I}|) + Q_{ai}Q_{\alpha I}\frac{f(|\mathbf{r}_{ai}-\mathbf{R}_{\alpha I}|)}{|\mathbf{r}_{ai}-\mathbf{R}_{\alpha I}|}\right] - \frac{1}{2}\sum_{i}^{N_{\mathrm{v}}}\sum_{a,b}^{N_{\mathrm{s}}}K_{ab}V_{ai}V_{bi} \tag{17}$$

where N_{v} is the number of solvent water molecules with N_{s} being the number of sites in a solvent molecule. The functions U and U' are the intramolecular potential functions for the peptide, eq 10, and solvent, respectively. $Q_{\alpha\mathrm{I}}$ includes the induced charge due to the intra- and intermolecular interactions. The electrostatic potential in the solvent reorganization energy term is written as

$$V_{ai} = \sum_{\beta\in I}^{N_I}\sum_{I}^{N_{\mathrm{u}}}Q_{\beta I}\frac{f(|\mathbf{R}_{\beta\mathrm{I}}-\mathbf{r}_{ai}|)}{|\mathbf{R}_{\beta\mathrm{I}}-\mathbf{r}_{ai}|} + \sum_{j(\neq i)}^{N_{\mathrm{v}}}\sum_{b}^{N_{\mathrm{s}}}Q_{bj}\frac{f(|\mathbf{r}_{bj}-\mathbf{r}_{ai}|)}{|\mathbf{r}_{bj}-\mathbf{r}_{ai}|} \tag{18}$$

where the first and second terms are the contributions from the solute and solvent, respectively.

The electrostatic parameters for a water molecule, that is, the atomic site charges and CRK matrix elements, were

calculated at the MP2 level with the 6-31G(2d,2p) basis set, and the results are $Q_O = -0.680$ and $Q_H = 0.340$ for the site charges and $K_{OO} = -4.760$, $K_{HH} = -2.122$, $K_{OH} = 2.380$, and $K_{HH'} = -0.2580$ in atomic units for the CRK matrix elements, respectively. The Lennard-Jones (LJ) parameters, ε and σ, were taken from the SPC model,[53] and the standard combination rule is applied for constructing the site−site interaction. The intramolecular motion of water is described by the Morse potential for the stretching mode and the harmonic potential for the bending one

$$U' = D_e[1 - e^{-a_e(r-r_{eq})}]^2 + D_e[1 - e^{-a_e(r'-r'_{eq})}]^2 + f_r(r - r_{eq})(r' - r'_{eq}) + \frac{1}{2}f_\theta r_{eq}^2(\theta - \theta_{eq})^2 \quad (19)$$

where $r_{eq} = 0.958$ Å and $\theta_{eq} = 102.9°$, the ab initio optimized values. The other parameters, D_e, a_e, f_r, and f_θ, are the same as in the previous study of water with the CRK model.[33]

For the solvent−solvent LJ terms in eq 17, the long-range part was cut off with the following function:

$$s(x) = \begin{cases} 1 & (r_{ij} < 0.9r_c) \\ 1 - 10x^3 + 15x^4 - 6x^5 & (0.9r_c < r_{ij} < r_c) \\ 0 & (r_c < r_{ij}) \end{cases} \quad (20)$$

where

$$x = (r_{ij} - 0.9r_c)/0.1r_c$$

Here, r_c is the cutoff length that is half of the box size and r_{ij} is the distance between the i and jth molecules. The electrostatic interactions between the solvent−solvent and solute−solvent are attenuated at a short range with the damping function given in eq 7, as in the case of the intramolecular interaction in the polypeptide. The Ewald sum method was employed to treat the long-range electrostatic interaction.

MD simulations of solvated alanine polypeptide systems were performed with one solute and 472 solvents for the tetrapeptide (471 solvents for the pentapeptide) in a cubic box with a length of 24.66 Å, and the periodic boundary condition was employed. The initial structures are conformer **1** in Table 2 for the tetrapeptide and the ab inito optimized structure for the pentapeptide (see Figure 4). The equations of motion were integrated using the Verlet algorithm with a time step of 0.5 fs. The charges on solute and solvent atomic sites are determined self-consistently at each time step. The equilibration of the system at 298 K was achieved by MD trajectory calculations for 30 ps with occasional temperature scaling, which was followed by a MD run of 100 ps with no temperature scaling.

The intensity of the IR spectrum was calculated by[54]

$$I(\omega) = \frac{\omega}{3\hbar\varepsilon_0 Vc}\tanh\left(\frac{\beta\hbar\omega}{2}\right)\int_{-\infty}^{\infty} dt\, e^{-i\omega t}\langle \mathbf{M}(t)\cdot\mathbf{M}(0)\rangle \quad (21)$$

where V is the volume of system, c is the speed of light, ε_0 is the vacuum dielectric permittivity, and $\beta = 1/k_BT$ with k_B and T being the Boltzmann constant and temperature,

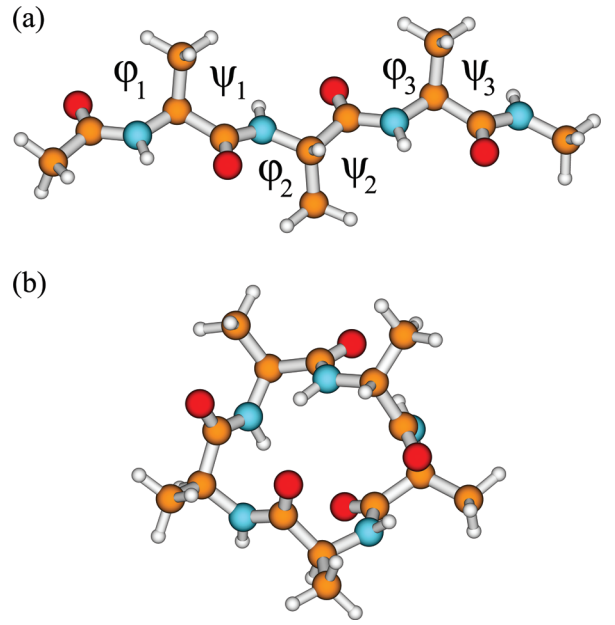

Figure 4. Starting conformations of tetrapeptide (a) and pentapeptide (b) in MD simulations. For tetrapeptide, the numbering of the dihedral angles is shown.

respectively. $\langle \mathbf{M}(t)\cdot\mathbf{M}(0)\rangle$ is the time correlation function (TCF) of the system dipole moment $\mathbf{M}$. Since we focused on the effect of charge polarization induced by the intra- and intermolecular interactions on the IR spectra, the contributions of the dipole moment of water solvent were subtracted from the system dipole moment in calculating the IR spectra.

To analyze the polarization effect on the spectrum, the dipole moment of the peptide in solution is decomposed into the gas-phase dipole moment $\mathbf{M}_G^U$ and the solvent-induced one $\mathbf{M}_I^U$ as

$$\mathbf{M}^U(t) = \mathbf{M}_G^U(t) + \mathbf{M}_I^U(t) \quad (22)$$

where the superscript U means the peptide dipole moment.[33] Note that the gas-phase dipole moment includes the effect of induced charge due to the intramolecular electrostatic interaction. The TCF is thus written as

$$C^U(t) = \langle \mathbf{M}^U(t)\cdot\mathbf{M}^U(0)\rangle = C_{GG}^U(t) + C_{II}^U(t) + C_{GI}^U(t) \quad (23)$$

where the first and second terms are the contributions from the gas phase and solvent-induced dipoles, respectively, and the third one is the cross term between them.

3.2. Results and Discussion. *Dihedral Angle Distribution.* Figure 5 shows the normalized distributions for the backbone dihedral angles (ϕ, ψ) of tetra- and pentapeptides, respectively. The dihedral angles for the main chain of tetrapeptide is numbered from the N terminus, as shown in Figure 4. For the tetrapeptide, the dihedral angle distributions show a large difference between the gas and solution phases despite the initial structures being the same. This is because the isolated tetrapeptide favors a compact form stabilized by the intramolecular hydrogen bonds, while the solvated peptide retains the initial extended form, which is stabilized

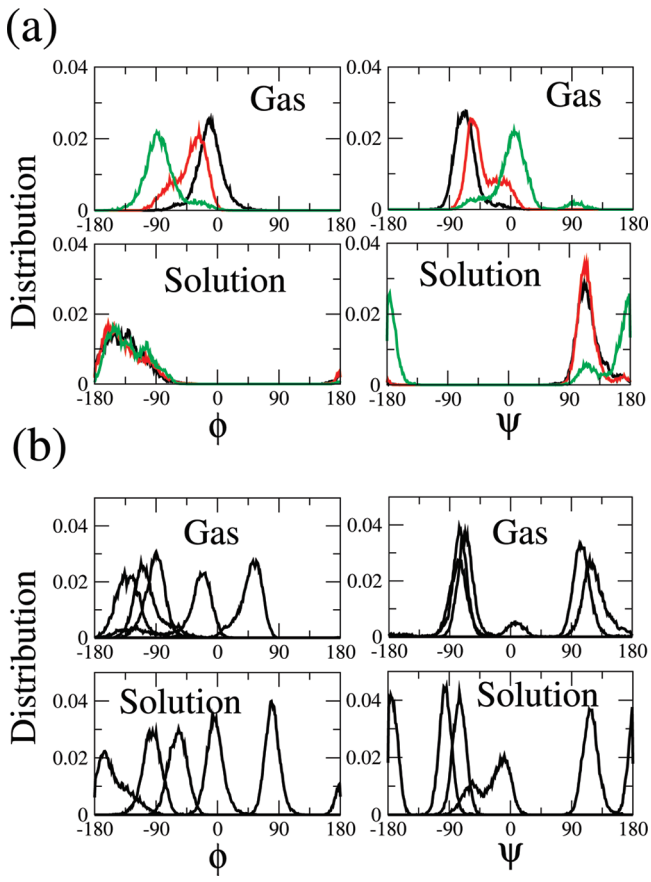

Figure 5. (a) Normalized distributions of backbone dihedral angles for tetrapeptide. Distributions of dihedral angles ϕ_1, ϕ_2, and ϕ_3 are written with black, red, and green, respectively. The right panel is the distributions of dihedral angles, ψ_1, ψ_2 and ψ_3. (b) Normalized distributions of dihedral angles for cyclic pentapeptide. The dihedral angles $\phi_1-\phi_5$ and $\psi_1-\psi_5$ are not distinguished.

by the hydrogen bondings with the surrounding water molecules. For the pentapeptide, there was no drastic conformational change within the present simulation time, 100 ps. In addition, a significant difference in distribution of the dihedral angles between the gas and solution is not observed, indicating that the dihedral angles of cyclic pentapeptide are rather restricted by taking the cyclic form.

Charge Fluctuation and Dipole Moment. The atomic site charges in the peptide fluctuate due to the thermal fluctuation of the electrostatic environment. We calculated the average site charges and their fluctuations for the four backbone atomic sites, N, H, C, and O, and the α- and β carbons, C_α and C_β. Similar analyses were performed by Liu et al.[56] for crambin with the QM/MM method, and for several small proteins by Patel and Brooks[17] with the fluctuating charge model. They showed that the polarization of atomic site charges is characterized by the chemical environments around the atoms due to the protein sequence. Since the alanine polypeptides were chosen as the solute in the present calculations, the effects of intra- and intermolecular hydrogen bonding on the site charges and their fluctuations can be discussed because the chemical environments around each backbone atom are very similar between the units. Here, we focused on the cyclic pentapeptide because the environment around all of the alanine units is equivalent by the symmetry. Figure 6 shows the average charges at the backbone atomic sites in each unit and their fluctuations obtained from the 100 ps trajectory records, where the left and right panels are those in the gas phase and in solution, respectively. It is noted that the site charge in each unit becomes the same if we take a large number of trajectories adequate to represent the thermal average. However, since we took one trajectory starting from the particular initial geometry and there was no conformational change during the 100 ps trajectory period, the site charge depends on the electrostatic field around the atom and has different values for different units, as seen in Figure 6. The average charge shows weak dependency on the amino acid unit in the gas phase, indicating that there is no strong correlation between the atomic charges and dihedral angle distributions. On the other hand, the difference in average charge becomes appreciable in the solution. For example, the site charge on N in unit 4 is larger than that in unit 1 by 0.5, implying that the solvated condition is different between the two atoms. As shown in Figure 6b, the charge fluctuation in solution is larger than that in the gas phase for all of the sites. It is also noted that the fluctuation at the β carbon C_β is approximately 1/10 of that at the α carbon C_α, because the large diagonal CRK element of C_β is canceled by the three equivalent large C_β−H off-diagonal elements under a nearly uniform electrostatic environment around the methyl group.

Figure 7 shows the distributions of the dipole moments of tetra- and pentapeptides in solution. The dipole moments at the peak position are almost the same, 11.4 D, for both the peptides. In order to see the role of charge polarization on the dipole moments, we calculated the dipole moments in the gas phase using the CRK charges as well as the reference ones. Note that, for calculating the dipole moments, we used the same trajectory records generated by the MD simulations in solution with the present polarizable force field. As seen in Figure 7, the dipole moment at the peak position increases by 3.3 D due to the intramolecular electrostatic interaction for the tetrapeptide. The solvent effect further enhanced the dipole moment by 1.7 D. For the cyclic pentapeptide, the increase of the dipole moment by the intramolecular interaction, 0.6 D, is smaller than that of the tetrapeptide. On the other hand, the solvent-induced contribution to the dipole moment is larger for pentapeptide, 3.1 D, than that of tetrapeptide. Compared with the pentapeptide, the distribution of the dipole moment of the tetrapeptide is broader. This is because the geometry of the tetrapeptide is flexible even in a solvent, while that of the pentapeptide is restricted to retain the cyclic form.

We further carried out ab initio electronic structure calculations for the tetra- and pentapeptides both in the gas and solution phases, and the resultant dipole moments were compared with those from the present CRK based model. For the calculations in solution, we employed the reference interaction site model self-consistent field (RISM-SCF) method, where the HF wave function with the 6-31G** basis set was used to describe the solute electronic structures, and the RISM integral equation was solved with the standard model of the water molecule.[55] From the MD trajectory

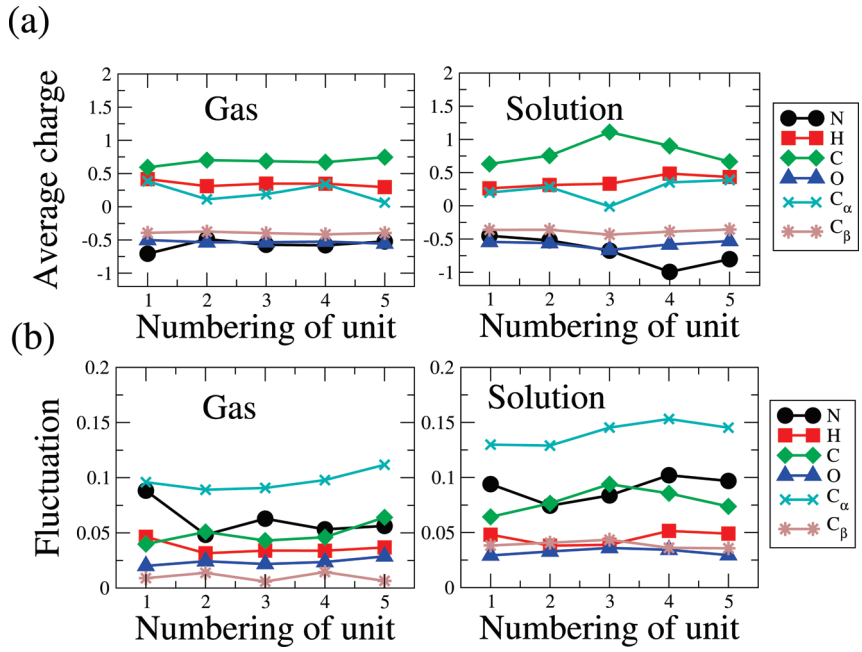

Figure 6. Average charges (a) and fluctuations (b) for backbone sites, N, H, C, and O, and α and β carbons, C_α and C_β, for cyclic pentapeptide in the gas phase (left) and solution (right). The *x* axis represents the numbering of the amino acid unit.

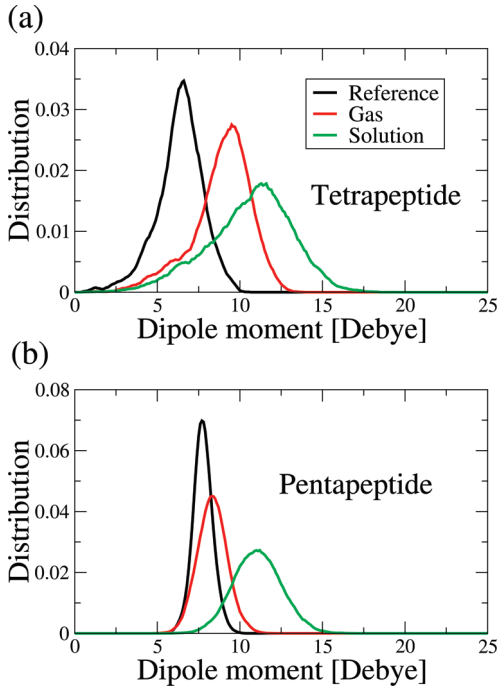

Figure 7. Normalized distributions of dipole moment for tetrapeptide (a) and pentapeptide (b) in solution. Black, red, and green correspond to reference, intramolecular induced, and intramolecular induced plus intermolecular induced dipole distributions, respectively.

records, we picked up one geometry for each peptide whose dipole moment is close to that of the peak position in Figure 7. The ab initio dipole moments for the tetrapeptide were 8.6 and 9.2 D in the gas and solution phases, respectively. The corresponding CRK values are 9.7 and 11.4 D, respectively. For the cyclic pentapeptide, the ab initio dipole moments were 9.3 and 11.3 D in the gas phase and solution, respectively, which are close to the CRK values of 8.3 and 11.4 D. These results indicate that the present model provides a reasonable estimate of charge polarization induced by the intra- and intermolecular hydrogen bondings. It is noteworthy that the dipole moments using the original AMBER99 charge set are 6.4 and 7.7 D for the tetra- and pentapeptides, respectively.

Solvation Structures. In order to examine the effect of peptide−water solvent hydrogen bondings on the charge distribution, we examined the solvation structures. We first calculated the radial distribution functions (RDFs) between the N, H, C, and O atoms in peptides and the water O atoms and then estimated the radius of the sphere determining the first solvation shell for 24 (tetra) and 30 (penta) atoms. The radii thus obtained are 4.5, 3.0, 4.0, 5.0, 4.0, and 5.0 Å for O, H, N, C, H(−C_α), and methyl C atoms, respectively. The corresponding values for pentapeptide are the same as those for the tetrapeptide. Next, we calculated the number of H_2O molecules by counting the numbers of H_2O in the first solvation spheres by checking a double counting of H_2O which is in more than two solvation spheres. The calculated number of H_2O in the first solvation shell was estimated to be 49−62 and 48−61 for the tetra- and peptapeptide, respectively. Since the cyclic pentapeptide has intramolecular hydrogen bondings, the number of coordinated H_2O becomes almost the same as that of the extended tetrapeptide in spite of the difference of the number of amino acid units between the two peptides.

Figure 8 compares the distribution of H_2O dipole moments in the first solvation shell of pentapeptide with that of the bulk water obtained by H_2O molecules far from the peptide atomic sites with a distance 10−11 Å. As seen in the figure, the average dipole moment of H_2O in the bulk water is 2.4 D, which is larger than the gas phase one, 1.9 D, obtained from the reference charges (see section 3.1). It is found that the dipole moment of H_2O around the N atoms is largely enhanced from the bulk one. On the other hand, the distributions of dipole moment around the other peptide

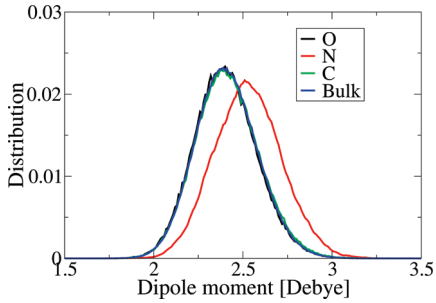

Figure 8. Distributions of water dipole moment in the first solvation shell of carbonyl oxygen (black), amide nitrogen (red), and carbon in the methyl group (green). The distribution in bulk water is also shown as a reference (blue).

atoms are very close to that of the bulk water. It is noted that the dipole moment distribution around the tetrapeptide was similar to that of the bulk solvent, though the dipole moment of the H_2O molecule forming a hydrogen bond with the NH or CO group is largely enhanced.

We further obtained the RDFs between the H (in NH)−O, O (in CO)−H, and C (in CH_3)−O site pairs to see the hydrogen bondings and hydrophobic interactions between the peptides and water. The dependence on the amino acid units was weak for the tetrapeptide because the extended form is favored in aqueous solution. For the cyclic pentapeptide, we observed a strong unit dependence of the number of hydrogen bondings, because the formation of the hydrogen bonding between the carbonyl O and water H atoms is prevented by the intramolecular hydrogen bonding. For the RDFs between the H atom in NH bond and water O atom in Figure 9b, we found two higher peaks in the first solvation shell. These peaks represent the strong hydrogen bonding in units 4 and 5, which is correlated to the enhancement of negative charges on the N atoms in units 4 and 5 in Figure 6. The sharp peaks observed in Figure 9c indicate the hydrohobic interaction of the methyl group with the water solvent.

IR Spectrum. Figure 10 shows the IR spectra up to 2000 cm^{-1} for the tetra- and pentapeptides, where the left and right panels are those in the gas and solution phases, respectively. The peak positions are almost the same between the two peptides in the gas phase. The two split bands at 1670 and 1770 cm^{-1} are attributable to the amide I mode, which arises mainly from the C═O stretching vibrations. Such a band splitting is not observed for the spectrum of N-methylacetamide (NMA) in the gas phase[43,45] because it is attributed to the coupling of the C═O stretching motions between the different units.[46,47,57,58] The split band at ∼1550 cm^{-1} is attributable to the amide II mode, which is mainly described by the out-of-phase combination of the N−H in-plane bend and C−N stretching vibrations. The amide III band emerges in the region 1200−1400 cm^{-1} and is rather complex, as seen in Figure 10. Compared with the amide I band, the intensities of amide II and III bands are too small. This is because the dependence of atomic charges on the nuclear coordinates[29] is not included in the present model and thus the change of dipole moment due to the stretching of the amide bond is largely underestimated.[43] The most prominent band around 790 cm^{-1} is the amide V band, which comes

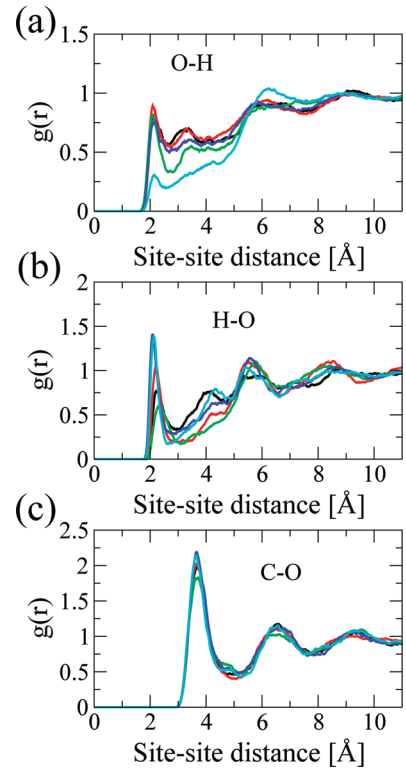

Figure 9. Radial distribution functions between (a) carbonyl oxygen and water H, (b) amide hydrogen and water O, and (c) methyl carbon and water O, respectively. The black, red, green, blue, and cyan curves correspond to the amide acid units from 1 to 5, respectively.

mainly from the out-plane N−H bending motion. The amide V band of pentapeptide shows splitting, indicating that the band profile is sensitive to the intra- and intermolecular environment. The neighbor band at 680 cm^{-1} is assigned to the amide IV and VI modes, which are composed of the out-of-plane C═O bend and the NCO deformation, respectively.

Compared with the gas phase, the shape of each band in aqueous solution becomes broader, and the intensity is largely enhanced. It is notable that the height of the amide V band of tetrapeptide does not increase by solvation, while that of pentapeptide becomes about twice that in the gas phase. In order to clarify the origin of such a difference in the amide V band shape between the tetra- and pentapeptide due to solvation, we decomposed the spectra into the three contributions, that is, the gas phase and solvent-induced self-terms and their cross term, eq 23. The results are displayed in the left panels of Figure 11, where we can see that the contribution of the solvent-induced self-term is small for the terapeptide, though such a term has the largest contribution for the pentapeptide. We further analyzed the vibrational density of state (VDOS) for the N−H bending motion for both the peptides by representing the correlation function of the N−H bond vector as the sum of two terms:

$$\langle \mathbf{R}(t)\cdot\mathbf{R}(0)\rangle/\langle \mathbf{R}(0)\cdot\mathbf{R}(0)\rangle = \sum_i \langle \mathbf{r}_i(t)\cdot\mathbf{r}_i(0)\rangle/\langle \mathbf{R}(0)\cdot\mathbf{R}(0)\rangle + \sum_i \sum_{j(\neq i)} \langle \mathbf{r}_i(t)\cdot\mathbf{r}_j(0)\rangle/\langle \mathbf{R}(0)\cdot\mathbf{R}(0)\rangle \quad (24)$$

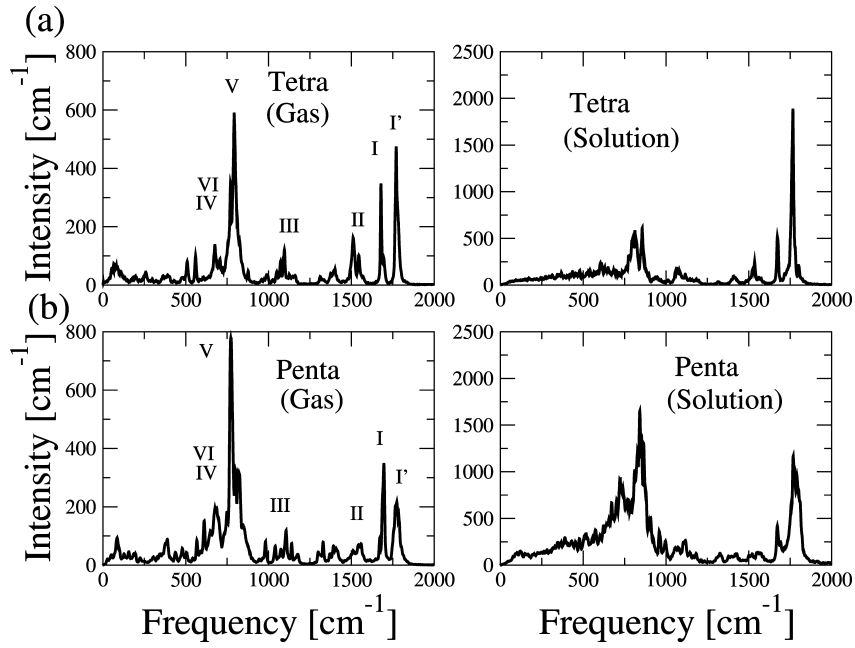

Figure 10. IR spectra for (a) tetrapeptide and (b) cyclic pentapeptide. Left and right panels are in the gas phase and in solution, respectively.

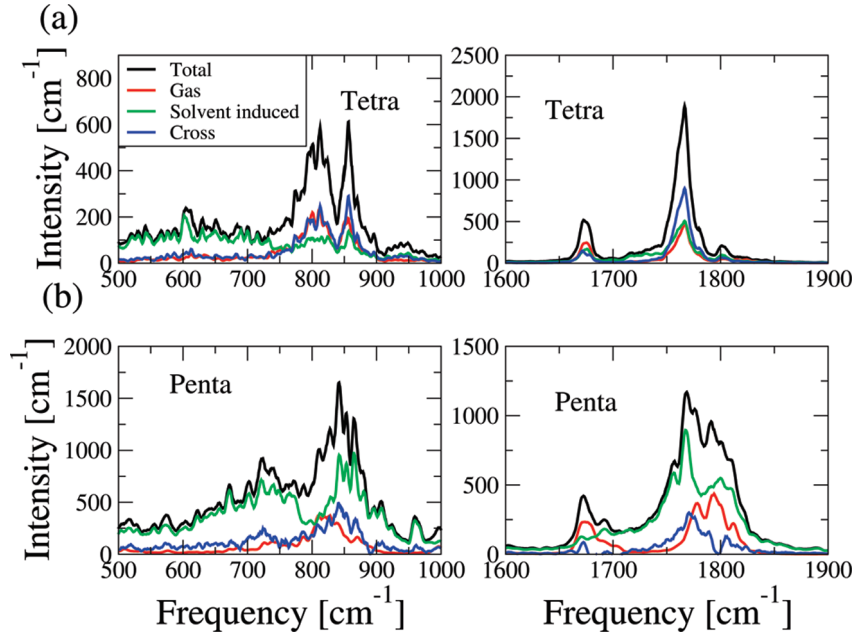

Figure 11. IR spectra of (a) tetra and (b) cyclic pentapeptide in solution in the regions of 500−1000 cm^{-1} (left) and 1600−1900 cm^{-1} (right). Decomposed into gas- and solvent-induced self-terms and their cross term.

where $\mathbf{r}_i$ is the N−H bond vector in the ith amino acid unit and $\mathbf{R}$ ($= \sum_i \mathbf{r}_i$) is the sum of N−H bond vectors. Thus, the VDOS spectrum is decomposed into the diagonal and off-diagonal contributions. The calculated VDOSs of tetra- and cyclic pentapeptides in the region 500−1000 cm^{-1} are shown in Figures S2 and S3 in the Supporting Information. Although both the diagonal and off-diagonal terms have positive intensities for the pentapeptide, the off-diagonal term has a negative contribution for the tetrapeptide, indicating that the enhancement of the amide V band for tetrapeptide is suppressed by the cancellation between the diagonal and off-diagonal terms.

The amide I band is known to be strongly affected by the dihedral structure of the backbone, and its frequency shift is utilized to probe the secondary structures of protein. The aqueous solvation effect on the band shape has been also discussed. As shown in Figure 10, the intensity is largely enhanced due to the solvation both for the tetra- and pentapeptides, and the increase of band intensity at 1770 cm^{-1} is notable. The right panels of Figure 11 show the components of IR spectra in the region 1600−1900 cm^{-1}, where we can see that the largest contribution to the intensity enhancement for the tetrapeptide is the cross term between the gas phase and solvent-induced ones. On the other hand,

the solvent-induced term is dominant, and the cross term is minor for the pentapeptide. One of the possible reasons for the small cross term of pentapeptide compared with the other two terms is that the directions of dipoles coming from the intramolecular and solvent-induced terms compensate each other. It is also noted that the band around 1770 cm^{-1} of pentapeptide shows a red shift from the gas-phase band due to the contribution from the solvent-induced self-term.

4. Conclusions

In the present paper, we developed a polarizable MM force field for polypeptides and proteins employing the CRK model. The effective charges and CRK matrices of polypeptides were constructed from those determined for the constituent amino acids. The electrostatic interaction potential thus obtained was combined with the AMBER99 force field. Furthermore, in order to reproduce the ab initio optimized geometries and relative energies of several conformers for 20 neutral and five charged dipeptides, we refitted the parameters of torsional potentials.

To evaluate the present MM force field, we applied it to MD simulations of the extended alanine tetrapeptide and cyclic pentapeptide. For the pentapeptide, the average site charges on the atoms in the main chain and their fluctuations in aqueous solution were compared with those in the gas phase in order to see the effect of intra- and intermolecular hydrogen bonding on the charge polarization. We also compared the dipole moments of the polypeptides by the present CRK-based model with those by ab initio calculations. As a result, the present model was shown to provide reasonable charge distributions both in the gas phase and in solution.

Finally, we calculated the IR spectra of alanine tetra- and pentapeptides in the gas phase and in solution in order to see the effect of intra- and intermolecular hydrogen bonding on the IR spectral profiles. For both of the peptides, split amide I modes were observed, which are attributed to the coupling of CO stretching modes between the different amino acid units. We also found that the intensity of the amide V mode for the tertapeptide does not increase due to the solvation, though that of pentapeptide is largely enhanced in solution.

It is noted that the present model should be regarded as the first step for the development of a polarizable protein force field employing the CRK method. In order to improve the model, we need to apply it to many problems such as solvation structure and free energy calculations and refine the parametrizations.

Acknowledgment. We are grateful to Professors A. Morita and S. Hayashi for their valuable discussions and comments. This work was supported by the Grant-in-Aid for Scientific Research from the Ministry of Education and Science, Japan.

Supporting Information Available: Figure S1 shows the schematics of structures for amino acids and terminal species considered in this work. The reference site charges for 20 neutral and five charged amino acids as well as terminal species are given in Tables S1−S27. The CRK matrix elements are also listed in Tables S1−S27. The fitted parameters for the torsional potentials are summarized in Table S28. The ab initio optimized values of key dihedral angles for stable conformers of dipeptides are summarized in Tables S29−S46. The rms deviations of relative energies and dihedral angles from the ab initio reference values are given in Tables S47−S63. The decomposition analyses of the VDOS spectra into the contributions from the diagonal and off-diagonal terms are shown in Figures S2 and S3. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) 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*, 765–784.

(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*, 5179–5197.

(3) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(4) Yin, D.; Mackerell, A. D. *J. Comput. Chem.* **1998**, *19*, 334–348.

(5) Feller, S. E.; Mackerell, A. D. *J. Phys. Chem. B* **2000**, *104*, 7510–7515.

(6) Klauda, J. B.; Brooks, B. R.; Mackerell, A. D.; Venable, R. M.; Pastor, R. W. *J. Phys. Chem. B* **2005**, *109*, 5300–5311.

(7) Daura, X.; Mark, A. E.; Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535–547.

(8) Oostenbrink, C.; Villa, A.; Mark, A. E.; Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(9) Schuler, L. D.; Daura, X.; Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(10) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638–6646.

(11) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077–13082.

(12) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(13) Dang, L. X.; Rice, J. E.; Caldwell, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 2481–2486.

(14) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1048–1057.

(15) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515–1531.

(16) Rick, S. W.; Stuart, S. J.; Bern, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(17) Patel, S.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1–15.

(18) Patel, S.; Mackerell, A. D.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504–1514.

(19) Lamoureux, G.; MacKerell, A. D.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(20) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

(21) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

(22) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(23) Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J. *J. Chem. Theory Comput.* **2005**, *1*, 169–180.

(24) Kim, B.; Young, T.; Hardrer, E.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2005**, *109*, 16529–16538.

(25) Soto, P.; Mark, A. E. *J. Phys. Chem. B* **2002**, *106*, 12830–12833.

(26) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, R. *J. Comput. Chem.* **2009**, *30*, 1701–1711.

(27) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *29*, 6290–6295.

(28) Morita, A.; Kato, S. *J. Am. Chem. Soc.* **1997**, *119*, 4021–4032.

(29) Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 89–100.

(30) Ishida, T.; Morita, A. *J. Chem. Phys.* **2006**, *125*, 074112.

(31) Isegawa, M.; Kato, S. *J. Chem. Phys.* **2007**, *127*, 244502.

(32) Morita, A.; Kato, S. *J. Chem. Phys.* **1998**, *108*, 6809–6818.

(33) Iuchi, S.; Morita, A.; Kato, S. *J. Phys. Chem. B* **2002**, *106*, 3466–3476.

(34) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(35) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, *27*, 781–790.

(36) Blatt, H. D.; Smith, P. E.; Pettitt, B. M. *J. Phys. Chem. B* **1997**, *101*, 7628–7634.

(37) Young, W. S.; Brooks, C. L., III. *J. Mol. Biol.* **1996**, *259*, 560–572.

(38) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605–612.

(39) Hu, H.; Elstner, M.; Hermans, J. *Proteins* **2003**, *50*, 451–463.

(40) Cui, Q.; Smith, H. *J. Chem. Phys.* **2002**, *118*, 279–290.

(41) Wang, Z.-X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699–1716.

(42) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182–3188.

(43) Gaigeot, M. P.; Vuilleumier, R.; Sprik, M.; Borgis, D. *J. Chem. Theory Comput.* **2005**, *1*, 772–789.

(44) Cho, M.; Yang, S. *J. Chem. Phys.* **2005**, *123*, 134503.

(45) Bouør, P. *J. Chem. Phys.* **2004**, *121*, 7545–7648.

(46) Moran, A. M.; Park, S.-M.; Mukamel, S. *J. Chem. Phys.* **2003**, *118*, 9971–9980.

(47) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253–11262.

(48) Torii, H. *J. Phys. Chem. B* **2007**, *111*, 5434–5444.

(49) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341–350.

(50) Ewing, C. S.; Walman, M.; Maple, J. R. *J. Phys. Chem. A* **2002**, *106*, 326–334.

(51) Boehm, H. J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129–7135.

(52) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.

(53) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F.; Hermans, J. *Intermolecular Force*; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.

(54) McQuarrie, D. A. *Statistical Mechanics*; Harper Collins Publishers: New York, 1976; pp 60−65.

(55) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1994**, *100*, 7443–7453.

(56) Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. *Proteins* **2001**, *44*, 484–489.

(57) Ganim, Z.; Chung, H. S.; Smith, A. W.; DeFlores, L. P.; Jones, K. C.; Tokmakoff, A. *Acc. Chem. Res.* **2008**, *31*, 432–441.

(58) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35*, 369–430.

CT900295U