THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Development of an Effective Polarizable Bond Method for Biomolecular Simulation

Xudong Xiao,[†,‡] Tong Zhu,[†] Chang G. Ji,*[,†,‡,§] and John Z. H. Zhang*[,†,§]

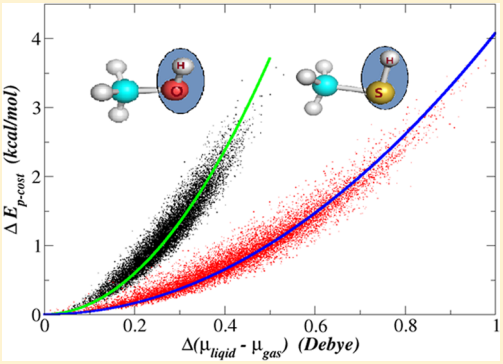
[†]State Key Laboratory of Precision Spectroscopy, Department of Physics, Institute of Theoretical and Computational Science, East China Normal University, Shanghai 200062, China

[‡]Institutes for Advanced Interdisciplinary Research, East China Normal University, Shanghai 200062, China

[§]NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

Supporting Information

ABSTRACT: An effective polarizable bond (EPB) model has been developed for computer simulation of proteins. In this partial polarizable approach, all polar groups of amino acids are treated as polarizable, and the relevant polarizable parameters were determined by fitting to quantum calculated electrostatic properties of these polar groups. Extensive numerical tests on a diverse set of proteins (including 1IEP, 1MWE, 1NLJ, 4COX, 1PGB, 1K4C, 1MHN, 1UBQ, 1IGD) showed that this EPB model is robust in MD simulation and can correctly describe the structure and dynamics of proteins (both soluble and membrane proteins). Comparison of the computed hydrogen bond properties and dynamics of proteins with experimental data and with results obtained from the nonpolarizable force field clearly demonstrated that EPB can produce results in much better agreement with experiment. The averaged deviation of the simulated backbone N−H order parameter of the B3 immunoglobobulin-binding domain of streptococcal protein G from experimental observation is 0.0811 and 0.0332 for Amber99SB and EPB, respectively. This new model inherited the effective character of the classic force field and the fluctuating feature of previous polarizable models. Different from other polarizable models, the polarization cost energy is implicitly included in the present method. As a result, the present method avoids the problem of over polarization and is numerically stable and efficient for dynamics simulation. Finally, compared to the traditional fixed AMBER charge model, the present method only adds about 5% additional computational time and is therefore highly efficient for practical applications.



I. INTRODUCTION

Extensive applications of molecular dynamics (MD) simulation to biomolecular systems rely on the use of an empirical molecular force field in which the interaction energy of biomolecules is represented as a function of atomic coordinates.[1−4] As a result, the correctness of the simulation result depends fundamentally on the accuracy of the force field employed in the simulation. Although great advances have been made in computational prediction of biomolecular properties, further improvement of force fields, especially by including proper polarization, is much needed.[5,6] A major challenge in the present force field development is how to accurately describe electrostatic interaction in biomolecular systems, in particular, how to properly include the polarization effect in MD simulation. Traditional fixed charge force field models are unable to adapt to environmental changes and therefore lack the polarization effect.[1,3]

For more than 30 years, many attempts have been made to explicitly include polarization effects in molecular modeling.[7−18] To date, there are several general models that include the polarization effect in the force field such as the fluctuating charge model,[10−12] Drude oscillator,[7−9] induced dipole,[13−17] electronic polarization via quantum mechanical treatment, or mixed QM/MM.[18] Reviews and accounts on the similarity and difference among these approaches are available.[19−21]

Electronic polarization results from redistribution of electrons when a molecule is placed in an electric field. There are two opposing energetic effects that occur during the polarization process. On one hand, polarization will enhance the interaction energy between the molecule and the external electric field in order to lower the energy of the system. On the other hand, the internal energy of the molecule will rise as a result of distortion of the electron charge distribution of the molecule. For example, when a polar molecule is placed in an electric field, these two opposing energetic effects will counter balance each other, and the molecular charge distribution will distort to a polarized state to establish a new system equilibrium (minimum energy state). On the basis of this simple physical picture, "equilibrium between polarization cost energy up and interaction energy down", we developed a new

Received: August 13, 2013
Revised: November 19, 2013
Published: November 19, 2013

   dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

fluctuating charge model for protein backbone polar groups. In this approach, the fluctuating charges are the sum of permanent (or reference) and induced charges that are determined by the difference of electrostatic potentials at different atoms. In particular, charges are only allowed to fluctuate between two oppositely charged atoms of the polar bond in this model. This polarizable model was applied to study site-specific backbone hydrogen bond energies in protein, and the theoretical predictions agreed very well with experimental measurement.[22]

In the current work, we extend this polarizable model to all polar groups of amino acids and develop a general "effective polarizable bond" model (EPB) for biomolecular simulation. In the present EPB model, energy cost during polarization is embraced into electrostatic interaction, and effective atomic charges fluctuate according to the local environment.

The rest of this article is organized as follows. Section II contains a detailed description of this method and tables of parameters. In Section III, the EPB model is applied to calculate the charge distribution of some proteins, and some results are presented. Conclusions are given in Section IV.

II. THEORETICAL METHOD

Effective Polarizable Bond. When a gas phase molecule is placed in an electric field such as in a polar solvent, the electron distribution of the molecule becomes distorted. The energy needed to distort the electron distribution of the gas phase molecule is called the distortion energy or polarization cost. The quantitative relationship between the polarization cost and the polarization state of certain chemical groups can be determined through electronic structure calculation.

The Schrodinger equation, which governs the state of the molecule's electrons in gas phase and in an electric field, is determined by quantum mechanical equations:

$$H_0\Psi_0 = E_0\Psi_0 \tag{1}$$

and

$$(H_0 + H')\Psi = E\Psi \tag{2}$$

where H_0 is the Hamiltonian of the molecule itself and H' is the Hamiltonian representing the interaction between the molecule and external electric field. The polarization cost energy of a polar group due to distortion of its electron distribution in the presence of an external electric field is defined by

$$\Delta E_{p-\cos t} = \langle\Psi|H_0|\Psi\rangle - \langle\Psi_0|H_0|\Psi_0\rangle \tag{3}$$

Figure 1 shows the relationship between polarization cost and the change of the dipole moment of the −SH and −OH group. The data in Figure 1 can be reasonably well-fit into a quadratic relation:

$$\Delta E_{p-\cos t} = k(\mu_{\text{liquid}} - \mu_{\text{gas}})^2 \tag{4}$$

$1/k$ represents polarizability of the relevant polar group. For example, Figure 1 shows that the −SH group is more polarizable than the −OH group.

Now, consider the transfer of a polar group −SH from gas phase into liquid phase, the energy of the system can be written as

$$E = E_{\text{self}} + E_{\text{ele}} = [k(\mu_{\text{liquid}} - \mu_{\text{gas}})^2] + [q_S\Phi_S + q_H\Phi_H] \tag{5}$$

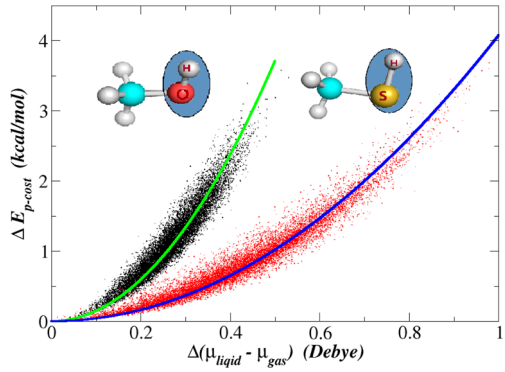


Figure 1. Polarization cost energy as a function of the change of the dipole moment. The black points represent data of the OH group which is fitted to a quadratic curve colored by green. The red points represent data of the SH group which is fitted to a quadratic curve colored by blue. For each group, 15 000 different configurations sampled from MD simulation were used in quantum mechanical calculation.

where q_S and q_H are, respectively, the ESP (electrostatic potential) charges of the S and H atoms of the −SH group. The Φ_S and Φ_H are electrostatic potentials at S and H atoms, respectively. In the present approach, the polarization can be treated as charge transfer between atoms of a polar group. If the amount of charge transfer from atom S to atom H is Δq, as in the −SH group, the final partial charges of the atoms are given by

$$q_S = q_S^{\text{gas}} + \Delta q \tag{6}$$

$$q_H = q_H^{\text{gas}} - \Delta q \tag{7}$$

where q_S^{gas} and q_H^{gas} are, respectively, the atomic charges of S and H atoms in gas phase (or reference charges). Thus, the change of dipole moment of the −SH group due to polarization (from gas phase to solvent) is given by

$$\Delta\mu = \mu_{\text{liquid}} - \mu_{\text{gas}} = \Delta q \cdot d_{SH} \tag{8}$$

where d_{SH} is the bond length of the S−H bond. Then, eq 5 can be rewritten as

$$E = k(\Delta q \cdot d_{SH})^2 + (q_S^{\text{gas}} + \Delta q)\Phi_S + (q_H^{\text{gas}} - \Delta q)\Phi_H \tag{9}$$

It should be noted that in the above approach, the direction of the induced dipole is along the direction of the polar bond. Thus, the approach is appropriate for treating polar groups with a single polar bond (or linearized polar bond).

Equation 9 gives the change of electrostatic energy of the molecule (the −SH group) in solvent. The equilibrium is reached by minimizing this energy with respect to variation of Δq

$$\frac{\partial E}{\partial \Delta q} = 0 \tag{10}$$

from which we obtain the charge transfer along the S−H bond under a given electrostatic potential,

$$\Delta q = \frac{(\Phi_H - \Phi_S)}{2{d_{SH}}^2 k} \tag{11}$$

Just as partial atomic charges can be introduced to describe electrostatic interaction, the effective charge concept can also

dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

be introduced in the fluctuating charge model. It is thus convenient to express the polarization cost term in eq 5 in a form of electrostatic interaction. For the SH group, eq 5 can be rewritten as:

$$\begin{aligned} E &= E_{self} + E_{ele} \\ &= k(\Delta q \cdot d_{S-H})^2 + [(q_S^{gas} + \Delta q)\Phi_S + (q_H^{gas} - \Delta q)\Phi_H] \\ &= \tilde{q}_S\Phi_S + \tilde{q}_H\Phi_H \end{aligned} \tag{12}$$

where $\tilde{q}_S$ and $\tilde{q}_H$are, respectively, effective fluctuating charges (EFQ) of S and H atoms. Combining eq 11 with eq 12, we obtain

$$\begin{aligned} E &= k(\Delta q \cdot d_{S-H})^2 + [(q_S^{gas} + \Delta q)\Phi_S + (q_H^{gas} - \Delta q)\Phi_H] \\ &= \Delta q \cdot (\Delta q k d_{S-H}{}^2) + [(q_S^{gas} + \Delta q)\Phi_S + (q_H^{gas} - \Delta q)\Phi_H] \\ &= \Delta q \cdot \left(\frac{(\Phi_H - \Phi_S)}{2d_{SH}{}^2 k} k d_{S-H}{}^2 \right) + [(q_S^{gas} + \Delta q)\Phi_S \\ &\quad + (q_H^{gas} - \Delta q)\Phi_H] \\ &= \frac{1}{2}\Delta q \cdot (\Phi_H - \Phi_S) + [(q_S^{gas} + \Delta q)\Phi_S + (q_H^{gas} - \Delta q) \\ &\quad \Phi_H] \\ &= \left[\left(q_S^{gas} + \frac{1}{2}\Delta q \right)\Phi_S + \left(q_H^{gas} - \frac{1}{2}\Delta q \right)\Phi_H \right] \\ &= \tilde{q}_S\Phi_S + \tilde{q}_H\Phi_H \end{aligned}$$

The effective fluctuating charges can be defined as:

$$\tilde{q}_S = q_S^{gas} + \Delta q/2 \tag{13}$$

$$\tilde{q}_H = q_H^{gas} - \Delta q/2 \tag{14}$$

which is essentially the result of linear response theory.[23]

The EFQ are directly used in MD simulation to correctly describe electrostatic interactions with polarization. The use of EFQ gives more balanced electrostatic interaction because it includes the effect of distortion or polarization cost in both energy and force. Electrostatic potential of an atom is determined by the charges of external potential and all other atoms except those in the same residue.[15]

Parameterization of the Fluctuating Charge Model. Since the energy expression in eq 5 or 9 is quite general, we can prefit the specific polarization parameters, such as reference atomic charges and polarizability k of all polar groups in amino acids. While previous work only treated polarization of backbone polar groups,[22] the current work extends fluctuating charge treatment to all polar groups including those of protein side chains, which is necessary for accurately modeling polarization of protein systems.

Fitting Protocol. Equations 3 and 4 can be used to determine the polarizable parameter k by quantum chemistry calculation for a given polar group such as the −SH bond. However, we notice that the polarizable parameter k is related to a linear polar bond that only have two atoms, thus we cannot directly parametrize k for polar groups that have more than two atoms such guanidine which has nine atoms. Here a new approach is employed in which ESP (electrostatic potential) charge is used to fit the polarizable parameter k.

First of all, we need to select appropriate molecular models (Figures S1−S11). Here, for the −SH bond, we choose CH_3SH because $-CH_3$ is a nonpolar group and simple enough. Also it best mimics the $-CH_2-$ group which usually links to polar groups in the side chain of amino acids.

To model under external electric field, the CH_3SH was solvated in an octahedron TIP3P water box and a 30 ns NPT MD simulation was performed to sample different configurations of water around CH_3SH molecule (CH_3SH is frozen). A total of 15,000 configurations were extracted from the trajectory. For each configuration, quantum chemistry calculation of CH_3SH was performed in electric field generated by the background charge of water molecules. MD simulation was performed using AMBER12 software[24] and quantum chemistry calculations were performed using DFT method (M06/6-31G**) in Gaussian09[25] (Figure 2).

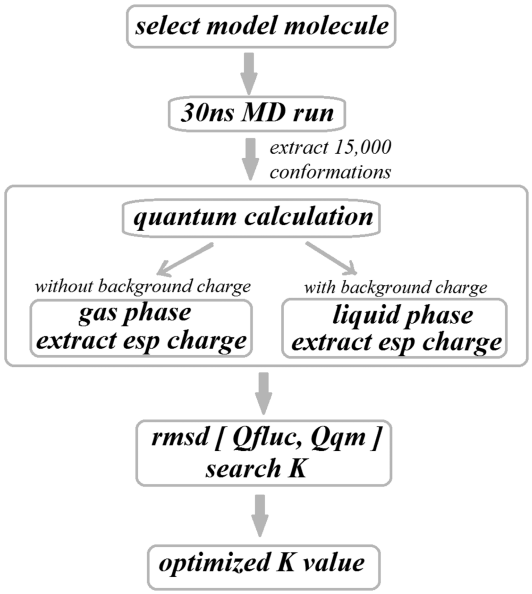


Figure 2. Flow diagram of the fitting procedure to obtain the value of polarizable parameter k.

The fluctuating charge of the polar atom obtained from eqs 6, 7, and 11 is thus compared to ESP charge directly computed from quantum chemistry calculation described above for the model molecular system (e.g., CH_3SH). For the entire set of 15,000 configurations, we can calculate the RMSD of the polar atom (root-mean-square deviation) between the fluctuating charge obtained from the EFQ model (Q_{fluc} from eq 6 or 7) and ESP charge from quantum chemistry calculation (Q_{qm}) given by the expression:

$$\begin{aligned} &\mathrm{RMSD}(Q_{fluc}, Q_{qm}) \\ &= \sqrt{\sum \left[q_H^{liquid} - \left(q_H^{gas} + \frac{\Phi_S - \Phi_H}{2k \cdot d_{S-H}{}^2} \right) \right]^2 / N} \end{aligned} \tag{15}$$

where d_{S-H} is the bond length of the S−H bond, $q_H{}^{liquid}$ and $q_H{}^{gas}$ is, respectively, the liquid phase and gas phase esp charge of the H atom, Φ_S and Φ_H are the electrostatic potentials at S and H atoms in liquid phase, and N is the number of configurations. The value of the polarizable parameter k is

dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

determined by minimizing RMSD$[Q_{fluc}, Q_{qm}]$.[26] We also fitted some parameters by using "gas phase interactions with electrostatic probes", and the results did not change.

For the more complicated polar groups, the fitting procedure was exactly the same as the −SH group described above, except that they were split into more fragments, each with a single polar bond.

List of Parameters. The list of fitted parameters of side chain polar groups in polar and charged amino acids including SER, THR, CYS, CYM, TYR, ASN, GLN, ASP, GLU, LYS, ARG, HIS, as well as the refined parameters of backbone polar groups, C═O and N−H are given in Table 1. The parameters

Table 1. Fitted EPB Parameters of the Side Chain and Backbone Polar Groups[a]

RES	bond	q_0		k	d
ASP	C^{γ}−O^{δ}	0.7406	−0.7720	3.28	1.26
ASH	C^{γ}−O^{δ}	0.6412	−0.6326	2.97	1.21
	O^{δ}−H^{δ}	−0.5365	0.4558	7.46	0.97
GLU	C^{δ}−O^{ε}	0.7258	−0.7790	3.28	1.26
GLH	C^{δ}−O^{ε}	0.6850	−0.5887	2.97	1.21
	O^{ε}−H^{ε}	−0.5990	0.4120	7.46	0.96
SER	O^{γ}−H^{γ}	−0.6301	0.4030	9.77	0.96
THR	O^{γ}−H^{γ}	−0.6689	0.4030	9.77	0.96
TYR	O^{η}−H^{η}	−0.5617	0.4030	8.41	0.96
CYS	S^{γ}−H^{γ}	−0.3296	0.2110	3.52	1.33
CYM	C^{β}−S^{γ}	−0.2000	−0.9257	1.73	1.81
ASN	C^{γ}−O^{δ}	0.7680	−0.6481	3.28	1.23
	N^{δ}−H^{δ}	−0.9681	0.4441	7.84	1.01
GLN	C^{δ}−O^{ε}	0.7513	−0.6648	3.28	1.23
	N^{ε}−H^{ε}	−0.9735	0.4415	7.84	1.01
LYS	N^{ζ}−H^{ζ}	−0.3778	0.3375	15.2	1.01
ARG	N^{ε}−H^{ε}	−0.4797	0.2958	8.75	1.01
	N^{η}−H^{η}	−0.8549	0.4439	7.14	1.01
HID	N^{δ}−H^{δ}	−0.3620	0.3458	8.76	1.01
	C^{ε}−N^{ε}	0.1610	−0.5280	3.27	1.31
HIE	N^{δ}−C^{ε}	−0.5210	0.1413	2.73	1.32
	N^{ε}−H^{ε}	−0.2840	0.3384	9.70	1.01
HIP	N^{δ}−$H^{\delta 1}$	0.4050	−0.1697	8.86	1.01
	C^{ε}−$H^{\varepsilon 1}$	0.2374	0.0137	7.62	1.08
	N^{ε}−$H^{\varepsilon 2}$	0.3520	−0.1336	10.8	1.01
	C^{δ}−$H^{\delta 2}$	0.2807	−0.1631	8.40	1.08
BB-I	C−O	0.5714	−0.5420	3.28	1.23
	N−H	−0.4478	0.3040	9.19	1.01
BB-II	C−O	0.4967	−0.5420	3.28	1.23
	N−H	−0.5267	0.3040	9.19	1.01
BB-III	C−O	0.6867	−0.5420	3.28	1.23
	N−H	−0.3772	0.3040	9.19	1.01

[a]The k is in kcal/mol·debye^{-2} and distance d in Å. BB- I, II, and III represent backbone polar groups for neutral, negative, and positive amino acids, respectively.

are divided into two classes, one is the set of gas phase charges and the other is the value of polarizable parameter k. The bond length used in the calculation and fitting is also given in the Table for reference purpose.

As mentioned above, for some complicated polar groups, they were split into several fragments. Each fragment is a polar bond containing two connected atoms, similar to simple polar groups (e.g., −SH) so the fitting approach is almost the same. There are exceptions such as guanidine for which the charge of the buried C atom is not allowed to fluctuate because the ESP charge was poor for buried atoms. In addition, for the side chain of unprotonated histidine, the atoms located in the two special polar bonds are allowed to fluctuate because they can be important in forming side chain H bonds. For backbone polar groups −CO and −NH, the *N*-methyl acetamide (NMA) is used as the model molecule to fit their parameters. All the parameters are listed in Table 1.

III. APPLICATIONS TO REALISTIC SYSTEMS

Several systems were studied here using the newly developed EPB model to investigate polarization effects in proteins. All initial structures were taken from the protein data bank. Most of the MD simulations were performed using the force fields of AMBER99SB[27] and AMBER99SB mixed with the EPB model. The modified version of AMBER12 package was used as a computational tool. Electrostatic potential on each atom was saved in the simulation for calculation of the fluctuating charge.

In the AMBER12 package, the electrostatic potential is calculated with the particle−mesh Ewald (PME) method. In the PME method, calculation of infinite summation of contribution from surrounding atoms is transformed from a conditionally convergent series into an absolutely convergent one by the introduction of a set of Gaussian screening charges added to and subtracted from the charge distribution. The electrostatic potential is broken into four parts in PME routines of AMBER12: (1) The "direct coulomb contribution", accumulated in the *short_ene_dip* subroutine in *short_ene.f*. (2) The "reciprocal space contribution", accumulated in the *grad_sum_dipolerc* subroutine in *ew_dipole_recip.f*. (3) The "adjustment and self corrections", accumulated in the *nb_adjust_dipole* subroutine of *ew_force.f*. (4) The "(1,4) corrections", accumulated in *do_14_dipole* of *extra_pts.f*. Changes to PME routines of AMBER12 were made in the code paths to save electrostatic potential. Fluctuating charges were calculated and updated in the main routine (*runmd.f*) according to the electrostatic potential on each atom.

The soluble proteins were embedded in a truncated octahedral TIP3P[28] water box, and counterions were added to neutralize the whole system. The membrane protein was embedded into 108 POPC lipid molecules, and a size of 68 × 70 × 96 box was used, the rest of which were filled by 8815 tip3p waters and 0.2 M KCl. The LIPID11 force field[29] was adopted for POPC molecule. Periodic boundary conditions and PME methods[30] were adopted to calculate the long-range electrostatic interactions. A cutoff of 10 Å was used for short-range interactions. Each system was relaxed in 10 000 steps with constraints on protein, followed by full minimization without any constraints. The system was then subjected to a 150 ps restrained MD simulation in NVT ensemble with protein atoms constrained at a force constant of 500 KJ/mol so that the solvent and ions can be properly positioned around the protein. Another 150 ps NPT simulation was run for adjusting the density of the system. Temperature was regulated using Langevin dynamics with the collision frequency setting to 2 ps^{-1}. All the MD simulations used a time step of 1 fs, and all the covalent bonds involving hydrogen atoms were constrained with the SHAKE algorithm.[31]

Range of Fluctuating Charge. *Fluctuation of Charges in Side Chains.* Partial charges on a particular residue are determined by their specific conformation and chemical environment due to other residues of the protein and solvents. For a specific amino acid, its polarization state when buried inside a protein can be very different from when it is located on

dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

a protein surface (Table S1). To illustrate this feature, four proteins (PDB codes: 1IEP, 1MWE, 1NLJ, 4COX) were studied. In the present calculation, a 3.5 ns NPT simulation is performed followed by a 1.5 ns production run for all four systems. The histogram of charge distribution for some side chain atoms are listed in Figure 3. We find that charges on the

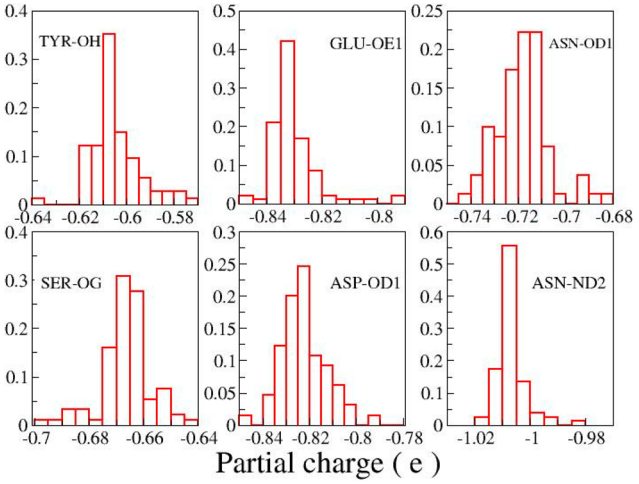


Figure 3. Partial charge distributions (percentage) for six types of side chain atoms. All side chains from four proteins (PDB code: 1IEP,1MWE,1NLJ,4COX) were counted.

same type of amino acid can differ significantly from each other. For example, the value of atomic charge ranges from −0.64 to −0.58 for OH@TYR, from −0.70 to −0.64 for OG@SER, and from −0.75 to −0.68 for OD@ASN in those four proteins we studied.

Fluctuation of Charges in Backbone. Figure 4 shows the average charges of the backbone hydrogen and oxygen for the

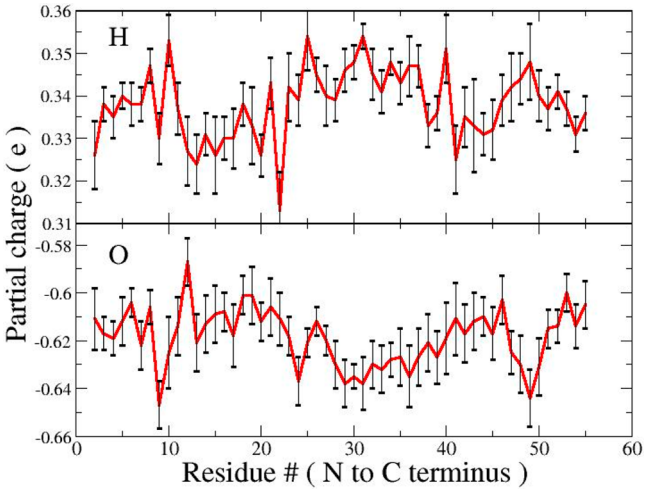


Figure 4. Fluctuation of backbone hydrogen and oxygen charges in 1PGB. Error bars represent the RMS fluctuation of charges during the entire MD simulation.

B1 binding domain of protein G[32] (PDB: 1PGB) as a function of position along the main chain from N to C terminus. The error bars measure the fluctuation of atomic charges during MD simulation (up to 10 ns here). The fluctuation behavior of the atomic partial charge has been investigated by other fluctuating charge models and QM/MM study. In the case of backbone oxygen atoms, the charges in CHARMM fluctuating charge force field[33] ranged from −0.8e to −0.9e, while the QM/MM study on crambin (PDB:1CRN) by Yang and co-workers[34] showed a range from −0.55e to −0.65e. For comparison, the range from the present EPB model is from −0.56e to −0.66e, which is in excellent good agreement with QM/MM study.

In recent years, with the increasing number of membrane protein crystal structures being solved,[35−38] the study of membrane protein becomes more and more popular. Since membrane proteins often span different regions with very different electrostatic environments, standard fixed charge force fields are not accurate to model different microenvironments of membrane protein effectively. So, the EPB model can be very effective to study membrane proteins.

The KcsA K+ channel(PDBID:1K4C) is a tetramer composed of four identical subunits and features the TVGYG sequence motif, which constructs the selectivity filter at an axis of the protein near the extracellular side.[35] KcsA provides very special microenvironments to test charges: the part which closely interacts with the lipid molecules is thought to be hydrophobic, and the channel part, which can conduct potassium ions selectively, is thought to be polar. The fluctuation of backbone oxygen partial charges in KcsA is plotted in Figure 5, and only one subunit's charges are shown, as the microenvironments of the four subunits are nearly identical.

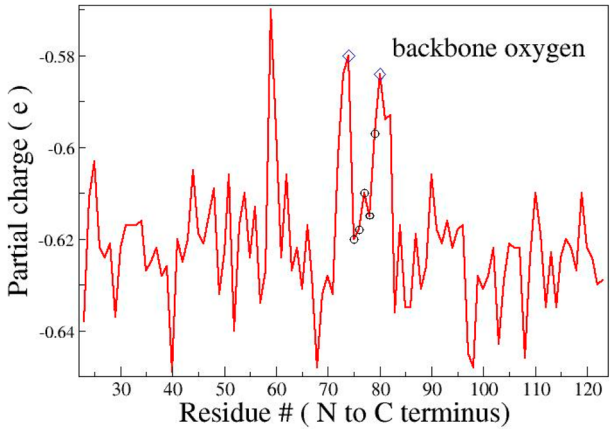


Figure 5. Fluctuation of backbone oxygen charges in KcsA membrane protein. The five circles in the middle denote the TVGYG sequence motif, and the two diamonds are the residues besides the TVGYG sequence motif.

Charges of carbonyl oxygen of the five filter residues range from −0.597 to −0.620, which is near the average value of all backbone carbonyl oxygen, implying that filter polar groups are not overly polarized compared to others. However, we found that charges on residues before and after the filter are relatively higher than others (see the two diamonds in Figure 5), which means that these backbone polar groups are less polarized than others. Detailed analysis of the dynamic structures of KcsA in MD simulation showed that the backbone CO groups of these residues are located in a hydrophobic environment without any hydrogen bonding partners near them. Statistical analysis of large amounts of protein structures showed that 95.4% of the backbone NH and CO of the completely buried peptide groups are hydrogen bonded.[39] Since a large amount of desolvation energy is needed to bury polar groups in the inner region of protein during the folding process, most of those polar groups

dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

prefer to form hydrogen bonds or salt bridges. The "unpaired" polar groups are in dynamically hot regions of proteins, and they prefer to find hydrogen bond partners. Residues before and after the TVGYG motif are dynamically unstable and can accelerate the ion passing process by destabilizing the filter in KcsA.[40] The dynamic character of those residues is important for protein to achieve its function (Figure 6).

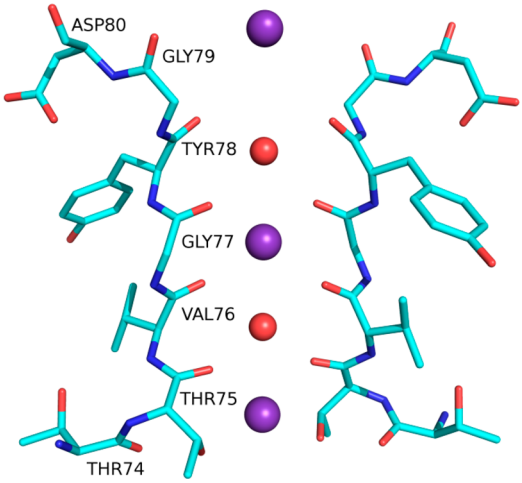


Figure 6. Selectivity filter of KcsA. Only two subunits are depicted for clarity. Blue spheres represent potassium ions, the red spheres represent water oxygen atoms, and the surrounding red sticks represent carbonyl oxygen atoms which are coordinated to potassium ions.

Conformation Changes in Proteins. Hydrogen bonds that are buried in the hydrophobic environment are relatively more stable than those on the protein surface which were exposed to solvents due to competition from water molecules. The dynamic stability of the hydrogen bond between NH@LYS4 and CO@LYS50 in 1PGB was examined here. The evolution of hydrogen bond length during MD simulation plotted in Figure 7 indicates that this hydrogen bond is broken in the 4 to 7 ns period. Detailed structural comparison around LYS4 before and after the breaking of the hydrogen bond was carried out, and Figure 8 shows that a water molecule was crowded into the hydrogen bond during MD simulation. The water molecule serves as a bridge between LYS4 and LYS50.

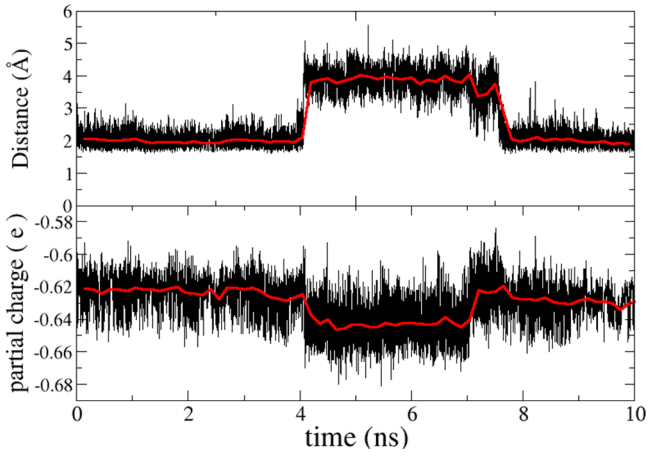


Figure 7. Distance between Lys50 backbone oxygen and Lys4 backbone hydrogen (top) and fluctuating charges on Lys50 backbone oxygen plotted as a function of simulation time (bottom). Average values are shown in red lines.

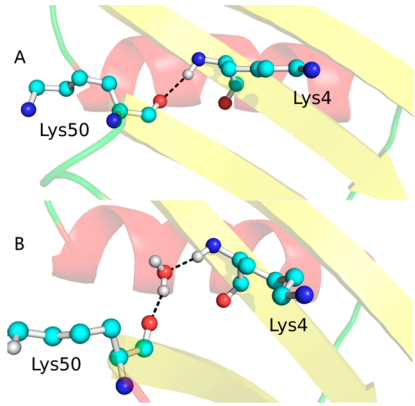


Figure 8. Structure of 1PGB around the Lys50−Lys4 hydrogen bond region. (A) Represents the state of forming a hydrogen bond, and (B) represents the state of forming a water bridge.

It is informative to investigate how polarization state was changed in such a short conformational process. Figure 7b (bottom) plots charge fluctuation of the LYS50s carbonyl oxygen. Before the breaking of the hydrogen bond, the average charge of oxygen was steady at about −0.62e. After the breaking of this hydrogen bond, the oxygen atom is more negatively charged (−0.64e) because it formed a water bridge, as shown in Figure 8B. After 7 ns, the lys50−lys4 hydrogen bond was formed again, and the average oxygen charge returned to the value around −0.62e.

Proteins in solvent may exist as an ensemble of multiple conformational states. Large scale conformational change is important for protein to achieve its biological function, including enzyme catalytic cycle and protein−ligand binding. Electrostatic environment may change during the conformational change. The preset EPB model can effectively capture the polarization changes associated with these conformational changes and thus make MD simulation more accurate and efficient.

Distribution of H−O Hydrogen Bond Distance. The hydrogen bond is a critical element in modulating the protein's three-dimensional structures. Electrostatic interaction is the main contributor of hydrogen bond interaction. To test whether the new charge model was suitable in describing the intraprotein hydrogen bond, hydrogen bond stability in MD simulation was analyzed for three soluble proteins: the SMN Tudor domain (PDBID:1MHN), the B1 immunoglobulin-binding domain of protein G(PDBID: 1PGB), and ubiquitin-(PDBID:1UBQ).

Distribution of the average length of all hydrogen bonds for each snapshot from the MD was pictured in Figure 9. As shown in Figure 9, for all three proteins, the vertical lines indicate average length of H−O bonds in the crystal structure, the maximum of the distributions under our EPB model were located at 1.93, 1.98, and 1.96, respectively; in contrast, under the pure AMBER99SB force field those values were located at 2.01, 2.08, and 2.07, respectively. This result indicates that the hydrogen bond structure was well-preserved in MD simulation when using effective fluctuating charge. The hydrogen bond length deviates a lot from the crystal structure when using

dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

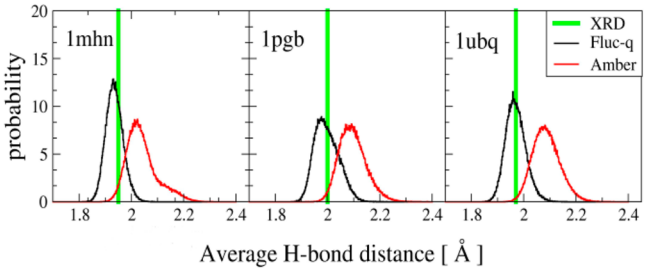


Figure 9. Distribution of H−O bond distances in MD simulation under our EFQ model and AMBER99SB for three proteins. Experimental values are indicted by XRD.

Amber99SB charge. Our findings suggest that polarization is important in stabilizing the protein's hydrogen bond.

Comparison of Simulated Protein Dynamics with NMR Experiments. Since NMR relaxation experiments provide direct observation of dynamic behavior of proteins, validation of the force fields can be accomplished by comparison of protein motions derived from MD simulation and NMR experiments.[27,41] The B3 immunoglobobulin-binding domain of streptococcal protein G (GB3), which is extensively studied by both NMR experiments[41] and computer simulations,[42]was used in our study. Here, we compare N−H bond order parameters of GB3 derived, respectively, from NMR experiment and MD simulations.

For this application, crystal structure of GB3 (PDB ID: 1IGD) is used in our simulation. Pretreatment of the structure is done according to NMR experiments as described in ref 43. After heating and equilibration, another 2 ns equilibration run is done at 297 K (NPT). A total of 15 structures are extracted from this 2 ns run, and they are used as the starting structures for the following 10 ns production runs (NVE). The simulation is done by using EPB and AMBER charge, respectively. The generalized order parameters are calculated from MD simulation as an ensemble average according to[43]

$$S^2 = \varsigma/2 < 3(\mu_i \cdot \mu_j)^2 - 1>$$

where μ_i and μ_j are the particular N−H bond vector scaled to unit magnitude at time moments of i and j, which are ensemble averaged. The order parameters are scaled by $\varsigma = 0.89$ since SHAKE is used to constrain chemical bonds involving hydrogen.[44]

Comparison of the simulated order parameters and the experimental results is shown in Figure 10. Each of these computed results is averaged from 15 production runs. The correlations between experimental and calculated order parameters are 0.83 and 0.86 for AMBER99SB charge and EPB charge, respectively. Since protein structure did not change much during simulation, the calculated order parameters all correlate well with the experimental values. From Figure 10, we can see that the EPB performs much better in reproducing dynamic behavior of the protein as measured by the NMR relaxation experiment. The averaged deviation of δS^2 is 0.0811 and 0.0332 for Amber99SB and EPB, respectively. The EPB gives the order parameters that are in much better agreement with the NMR experiment. A similar phenomenon was observed by Palmer et al;[42] that is, many order parameters derived from simulation with the Amber99SB, Amber03, and OPLS-AA force field all gave much lower values than those from the NMR experiments. The lower values of order parameters indicate overflexibility of the protein and demonstrate that the standard force field may "allow too much motion".[43] Without polarization, the conformational ensemble sampled from MD simulation may deviate substantially from the real native state. Our simulation and previous study suggest that systematic errors exist in MD simulation with a traditional fixed charge model due to the lack of a polarization effect.

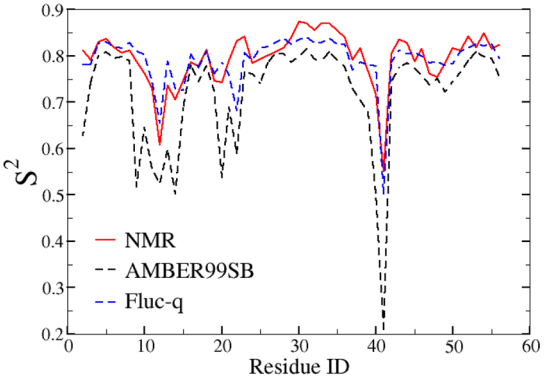


Figure 10. Comparison of NMR spin relaxation order parameters calculated from simulations under EFQ and Amber99SB with experimental measurements.

These results present strong evidence that it is important to include polarization effect in modeling protein's structure and dynamics. Including polarization is critical to get creditable predictions from molecular modeling. The EPB properly describes the polarization effect in proteins dynamically. The EPB model keeps the "effective charge" character of the classical force fields and provides a good correction to the current force field for MD simulation of protein through introducing "fluctuating" character.

IV.. CONCLUSIONS

We presented a partial polarizable EPB model for exploring protein dynamics in MD simulation. The parameters were derived from large scale quantum mechanical calculation of model molecules under different electrostatic environments. In this EPB model, the polarization cost or distortion energy in polarization is effectively included by using effective fluctuating atomic charges. This feature differs from other polarizable models. In previous polarizable models,[7−18] although polarization cost is explicitly included in energy function, the effect is not included in the calculation of force during MD simulation.[7−18] In reality, the polarization cost prevents two oppositely charged groups, such as NH and CO, from approaching too close to each other.

We have employed our EPB model to simulate different systems, both soluble protein and membrane protein. Protein structure and dynamics were better described in the simulation using the fluctuating charge than using traditional AMBER99SB charge. We did not add too much computational complexity during simulation because gathering the electrostatic potential on each atom does not cost much time. Thus, compared to the traditional fixed AMBER charge model, our new method only adds about 5% additional CPU time.

Our new charge method can also be used in Monte Carlo simulations. Like in the Car−Parrinello molecular dynamics approach, the effective fluctuating charges can be treated as additional degrees of freedom with a specific temperature T_{efq} in the Monte Carlo simulation.[45] The movement of charges can

dx.doi.org/10.1021/jp4080866 | J. Phys. Chem. B 2013, 117, 14885−14893

be accepted/rejected using the standard metropolis acceptance rule with T_{efq}.

It should be noted that all other parameters, except the charge, were obtained from the Amber99SB force field. Further work on optimizing the initial charges and Lennard–Jones parameters is ongoing in our lab to further improve the EPB method.

■ ASSOCIATED CONTENT

Supporting Information

Partial charges on the protein-protein binding interface upon binding calculated for the Barnase–barstar complex by our EFQ model and molecular models of pertinent compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: chicago.ji@gmail.com (C.G.J.).
*E-mail: john.zhang@nyu.edu (J.Z.H.Z.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the National Natural Science Foundation of China (Grants No. 21003048, 10974054, and 20933002). C.G.J. is also supported by the "Fundamental Research Funds for the Central Universities" and the Open Research Fund of the State Key Laboratory of Precision spectroscopy, East China Normal University. C.G.J. thanks Dr. John D. Chodera for his generous help on modifying the Amber source code. We also thank the Computer Center of ECNU for providing us computational time.

■ REFERENCES

(1) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(2) MacKerell, A. D.; Bashford, D.; Bellott, M.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(3) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(4) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*, 1205–1218.
(5) Freddolino, P. L.; Harrison, C. B.; Liu, Y. X. Challenges in protein-folding simulations. *Nat. Phys* **2010**, *6*, 751–758.
(6) Tong, Y.; Mei, Y.; Li, Y. L.; et al. Electrostatic polarization makes a substantial contribution to the free energy of avidin–biotin binding. *J. Am. Chem. Soc.* **2010**, *132*, 5137–5142.
(7) Kunz, A. P. E.; van Gunsteren, W. F. Development of a nonlinear classical polarization model for liquid water and aqueous solutions: COS/D. *J. Phys. Chem. A* **2009**, *113*, 11570–11579.
(8) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185–5197.
(9) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; et al. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245–249.
(10) Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1–15.
(11) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; et al. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23*, 1515–1531.
(12) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; et al. Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A* **2004**, *108*, 621–627.
(13) Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and *N*-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.* **2001**, *22*, 1048–1057.
(14) Wang, Z. X.; Zhang, W.; Wu, C.; et al. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781–790.
(15) Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. *J. Chem. Theory Comput.* **2007**, *3*, 1987–1992.
(16) Wang, J. M.; Cieplak, P.; Li, J.; et al. Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091–3099.
(17) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; et al. Current Status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
(18) Friesner, R. A. Modeling polarization in proteins and protein–ligand complexes: methods and preliminary results. *Adv. Protein Chem.* **2006**, *72*, 79–104.
(19) Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; et al. Polarization effects in molecular mechanical force fields. *J. Phys.:Condens. Matter* **2009**, *21*, 333102.
(20) Yu, H. B.; van Gunsteren, W. F. Accounting for polarization in molecular simulation. *Comput. Phys. Commun.* **2005**, *172*, 69–85.
(21) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11–28.
(22) Ji, C. G.; Xiao, X. D.; Zhang, J. Z. H. Studying the effect of site-specific hydrophobicity and polarization on hydrogen bond energy of protein using a polarizable method. *J. Chem. Theory Comput.* **2012**, *8*, 2157–2164.
(23) Li, Y.; Ji, C. G.; Xu, W. X.; et al. Dynamical stability and assembly cooperativity of beta-sheet amyloid oligomers—effect of polarization. *J. Phys. Chem. B* **2012**, *116*, 13368–13373.
(24) Case, D. A.; Darden, T. A.; Cheatham, T. E. et al. *AMBER 12*; University of California: San Francisco, 2012.
(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; et al.*Gaussian 09*, revision B.01; Gaussian, Inc.: Wallingford, CT, 2009
(26) Rick, S. W.; Berne, B. J. Dynamical fluctuating charge force fields: the aqueous solvation of amides. *J. Am. Chem. Soc.* **1996**, *118*, 672–679.
(27) Hornak, V.; Abel, R.; Okur, A.; et al. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.
(28) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.
(29) Skjevik, A. A.; Madej, B. D.; Walker, R. C.; Tiegen, K. LIPID11: a modular framework for lipid simulations using amber. *J. Phys. Chem. B* **2012**, *116*, 11124–11136.
(30) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(31) Andersen, H. C. RATTLE: a "velocity" version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(32) Gallagher, T.; Alexander, P.; Bryan, P.; et al. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **1994**, *33*, 4721−4729.
(33) Patel, S.; Mackerell, A. D.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Phys.* **2004**, *25*, 1504−1514.
(34) Liu, H. Y.; Elstner, M.; Kaxiras, E.; et al. Quantum mechanics simulation of protein dynamics on long timescale. *Proteins* **2001**, *44*, 484−489.
(35) Zhou, Y. F.; Morais-Cabral, J. H.; Kaufman, A.; et al. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 angstrom resolution. *Nature* **2001**, *414*, 43−48.
(36) Long, S. B.; Campbell, E. B.; MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science* **2005**, *309*, 897−903.
(37) Long, S. B.; Tao, X.; Campbell, E. B.; et al. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. *Nature* **2007**, *450*, 376−373.
(38) ter Haar, E.; Koth, C. M.; Abdul-Manan, N.; et al. Crystal structure of the ectodomain complex of the CGRP receptor, a class-B GPCR, reveals the site of drug antagonism. *Structure* **2010**, *18*, 1083−1093.
(39) Schell, D.; Tsai, J.; Scholtz, J. M.; et al. Hydrogen bonding increases packing density in the protein interior. *Proteins* **2006**, *63*, 278−282.
(40) Noskov, S. Y.; Berneche, S.; Roux, B. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* **2004**, *431*, 830−834.
(41) Hall, J. B.; Fushman, D. Variability of the N-15 chemical shielding tensors in the B3 domain of protein G from N-15 relaxation measurements at several fields. Implications for backbone order parameters. *J. Am. Chem. Soc.* **2006**, *128*, 7855−7870.
(42) Trbovic, N.; Kim, B.; Friesner, R. A.; et al. Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins* **2008**, *71*, 684−694.
(43) Case, D. A. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* **2002**, *35*, 325−331.
(44) Case, D. A. Calculations of NMR dipolar coupling strengths in model peptides. *J. Biomol. NMR* **1999**, *15*, 95−102.
(45) Martin, M. G.; Chen, B.; Siepmann, J. I. A novel Monte Carlo algorithm for polarizable force fields: application to a fluctuating charge model for water. *J. Chem. Phys.* **1998**, *108*, 3383−3385.