THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

Development of Polarizable Models for Molecular Mechanical Calculations I: Parameterization of Atomic Polarizability

Junmei Wang,[†] Piotr Cieplak,[‡] Jie Li,[§] Tingjun Hou,[∥] Ray Luo,[⊥] and Yong Duan*[,§]

[†]Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390, United States

[‡]Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, United States

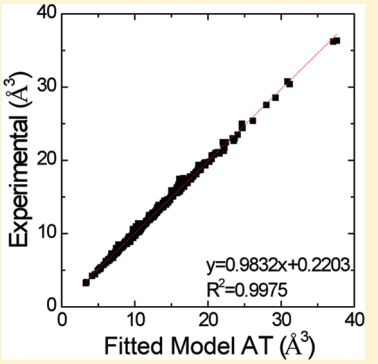
[§]UC Davis Genome Center and Department of Applied Science, University of California, Davis, One Shields Avenue, Davis, California 95616, United States

[∥]Functional Nano & Soft Materials Laboratory (FUNSOM), Soochow University, Suzhou 215123, China

[⊥]Department of Molecular Biology and Biochemistry, Department of Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States

Supporting Information

ABSTRACT: In this work, four types of polarizable models have been developed for calculating interactions between atomic charges and induced point dipoles. These include the Applequist, Thole linear, Thole exponential model, and the Thole Tinker-like. The polarizability models have been optimized to reproduce the experimental static molecular polarizabilities obtained from the molecular refraction measurements on a set of 420 molecules reported by Bosque and Sales. We grouped the models into five sets depending on the interaction types, that is, whether the interactions of two atoms that form the bond, bond angle, and dihedral angle are turned off or scaled down. When 1−2 (bonded) and 1−3 (separated by two bonds) interactions are turned off, 1−4 (separated by three bonds) interactions are scaled down, or both, all models including the Applequist model achieved similar performance: the average percentage error (APE) ranges from 1.15 to 1.23%, and the average unsigned error (AUE) ranges from 0.143 to 0.158 Å^3. When the short-range 1−2, 1−3, and full 1−4 terms are taken into account (set D models), the APE ranges from 1.30 to 1.58% for the three Thole models, whereas the Applequist model (DA) has a significantly larger APE (3.82%). The AUE ranges from 0.166 to 0.196 Å^3 for the three Thole models, compared with 0.446 Å^3 for the Applequist model. Further assessment using the 70-molecule van Duijnen and Swart data set clearly showed that the developed models are both accurate and highly transferable and are in fact have smaller errors than the models developed using this particular data set (set E models). The fact that A, B, and C model sets are notably more accurate than both D and E model sets strongly suggests that the inclusion of 1−2 and 1−3 interactions reduces the transferability and accuracy.



1. INTRODUCTION

With the continuing growth in computer power, the ability to model long-time scale events and large scale conformational transitions by molecular mechanical (MM) simulations has increased significantly. These large scale conformational changes can potentially lead to substantial variations in the dielectric environment. Thus, a widely recognized critical need in the field of molecular mechanical modeling is a model that can accurately represent the polarization effect.

A variety of models have been explored to represent the polarization effect in molecular systems. In AMBER, the polarization is modeled by the induced atomic dipoles.[1] Berne−Friesner groups utilized a fluctuating charge model in the context of OPLS-AA.[2] Brooks−Mackerell−Roux investigated the various fluctuating charge and drude oscillator models in CHARMM.[3] Ponder and coworkers explored a multipole expansion scheme and point polarizabilities in their AMOEBA force field.[4,5] Other methods include electronic polarization via quantum mechanical (QM) treatment,[6] mixed QM/MM,[7,8] polarization treatment using a continuum dielectric,[9,10] and so on. Recently, a review on polarization effects in molecular mechanical force fields was presented by Cieplak et al.[11]

The induced dipole model is the most studied approach for molecular polarization. It is employed by several force fields, which include OPLS/PFF, AMOEBA, and AMBER (ff02EP[12] and ff02r1[13]). The induced dipole interaction model implemented in the AMBER polarizable force fields represents the polarization effects at the atomic level, and the induced dipoles are

Received: December 21, 2010
Revised: February 1, 2011
Published: March 10, 2011

calculated using the Applequist scheme.[14] A potential difficulty of this scheme arises when two induced dipoles are too close that may lead to a polarization catastrophe.[5] Thole proposed a scheme to alleviate this deficiency by attenuating the dipole field tensor T_{pq}[15,16] that varies with the distance, giving rise to the Thole models. In AMOEBA, Ponder and Ren expanded the Thole scheme by including the interaction between induced dipoles and higher permanent electric moments (up to quadrupoles),[4,5] which can be derived from a distributed multipole analysis.[17]

In the dipole interaction models, the magnitude of the induced dipole moment μ_p at atom p is proportional to its atomic polarizability α_p. Because atomic polarizability plays a pivotal role in polarization calculations, accurate polarizability parameters are essential in the development of polarizable force field. The atomic polarizability parameters are obtained by fitting to either experimental or QM molecular polarizabilities or QM electrostatic potentials. In AMBER, for example, which utilizes the Applequist model and parameters, the parameter set was derived to reproduce the experimental molecular polarizabilities of 41 small molecules.[14] Dehez et al. derived atomic polarizabilities to reproduce the induction energies obtained by QM perturbation theory.[6] Kaminski et al. calculated a molecule's response to a dipolar probe located in a number of positions around the molecule using a density function theory; the perturbation of the electrostatic potential was then used to fit isotropic atomic polarizabilities.[18] Elking et al. proposed to replace atomic point charges with Gaussian charge densities, and the atomic polarizabilities were derived in a way similar to Kaminski's approach.[19] The calculated atomic polarizabilities have errors ranging from 1 to 5% depending on molecular species and methods.[11]

In this work, we present new isotropic atomic polarizabilities that reproduce high quality experimental molecular polarizabilities.[20] The experimental molecular polarizabilities were obtained by measuring the refractive index n, which is related to polarizability through the Lorentz−Lorenz equation (eq 1)

$$R = \left(\frac{n^2-1}{n^2+2}\right)\frac{M}{\rho} = \frac{4}{3}\pi N_a \alpha \qquad (1)$$

where α is the molecular polarizability, n is the refractive index, M and ρ are molecular weight and molar volume, respectively, and N_a is the Avogadro constant. The experimental polarizabilities determined by measurements of refractive index, are quite accurate, with a typical error of 0.5%. As an additive molecular property, it is possible to calculate the molecular polarizability by summing up the contributions from each element, atom type, or both.[20,21] Wang et al. recently published a set of empirical models, and the best one, which utilizes 14 atom types, achieves AUE, RMSE (root-mean-squares error), and APE of 0.147 Å^3, 0.219 Å^3, and 1.24%, respectively.[22]

The widely used data set for the atomic polarizability parametrization is the one used by van Duijnen and Swart[15] with a total of only 70 molecules (52 in the training set and 18 in the test set). The data set lacks iodides and phosphoric compounds that are commonly seen in biomolecular applications. Recently, Bosque and Sales reported a substantially larger data set of 420 diverse molecules.[20] Here taking advantage of the high-quality 420-molecule polarization data set of Bosque and Sales, we develop high-quality isotropic atomic polarizabilities for the Applequist's and three Thole's models by fitting to the experimental molecular polarizabilities. A genetic algorithm (GA) was used to optimize the average percent error. The initial validation is based on the comparison with experimental data of the van Duijnen and Swart[15] set of 70 molecules that were not included in the fitting. Further validation is provided in the companion paper of part II of the series.

2. DIPOLE INTERACTION MODELS

For a collection of N points, polarizable dipoles placed in a homogeneous electric field E, the induced dipole moment at point p (μ_p) is calculated by

$$\mu_p = \alpha_p[E_p - \sum_{q\neq p}^{N} T_{pq}\mu_q] \qquad (2)$$

where α_p is the atomic polarizability and $\boldsymbol{T}_{pq}$ is the dipole field tensor.

$$T_{pq} = \frac{f_e}{r_{pq}^3} I - \frac{3f_t}{r_{pq}^5}\begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z_2 \end{bmatrix} \qquad (3)$$

where $\boldsymbol{I}$ is a unit matrix; f_e and f_t are the distance-dependent screening functions. There are several widely used forms of f_e and f_t depending on the way the electron density is represented. In Applequist's model, $f_e = 1$ and $f_t = 1$.

It has been noted that Applequist's model may lead to infinite polarization by the cooperative interaction between two nearby inducible dipoles,[15,16,23] resulting in "polarization catastrophe". Thole[15,16] proposed solutions to this problem by introducing distance-dependent screening functions, f_e and f_t. In the linear screening function form, f_e and f_t are

$$\begin{aligned} v &= r_{pq}/[a(\alpha_p\alpha_q)^{1/6}] \\ \text{if } (v >= 1)\, f_e &= 1.0,\ f_t = 1.0 \\ \text{if } (v < 1)\, f_e &= 4v^3 - 3v^4,\ f_t = v^4 \end{aligned} \qquad (4)$$

In the exponential form, f_e and f_t are

$$\begin{aligned} v &= r_{pq}/[a(\alpha_p\alpha_q)^{1/6}] \\ f_e &= 1 - \left(\frac{v^2}{2} + v + 1\right)\exp(-v) \\ f_t &= 1 - \left(\frac{1}{6}v^3 + \frac{1}{2}v^2 + v + 1\right)\exp(-v) \end{aligned} \qquad (5)$$

In another exponential form adopted by Ren and Ponder,[5] f_e and f_t are determined by

$$\begin{aligned} v &= r_{pq}/[a(\alpha_p\alpha_q)^{1/6}] \\ f_e &= 1 - \exp(-v^3) \\ f_t &= 1 - (v^3+1)\exp(-v^3) \end{aligned} \qquad (6)$$

where α_p and α_q are the atomic polarizabilities of atoms p and q, respectively, a is the screening factor, and r_{pq} is the distance between atoms p and q.

For a molecule containing N atoms, eq 2 can be rewritten as $M = AF$,[15,24] where M is a $3N$-dimension vector containing the induced atomic dipole moments, F is a $3N$-dimension vector containing the electric field, and A is the $3N \times 3N$ atomic polarizability tensor, and the elements of the inversion matrix of

A take a form of $\delta_{ij}\alpha_i^{-1}+T_{ij}$

$$A^{-1} = \begin{bmatrix} \alpha_1^{-1} & T_{12} & \dots & T_{1N} \\ T_{22} & \alpha_2^{-1} & \dots & T_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N1} & T_{N2} & \dots & \alpha_N^{-1} \end{bmatrix} \quad (7)$$

The isotropic molecular polarizability is the trace of A. The $3N \times 3N$ polarizability tensor, A, can be reduced to a normal 3×3 molecular polarizability tensor B[15,24] by

$$B_{mn} = \sum_{i,j=1}^{N} (A_{ij})_{mn} \quad (8)$$

and the isotropic molecular polarizability is

$$\alpha_M = (B_{11} + B_{22} + B_{33})/3 \quad (9)$$

In our model, the molecular polarizabilities are not considered to be adjustable parameters but rather well-defined physical properties of molecules of interest. Therefore, we optimized the isotropic atomic polarizability parameters α to reproduce the experimental molecular polarizabilities. In this way, the atomic polarizability α has a clear physical meaning, rather than just being a parameter.

3. METHODS

In many polarizable force fields, especially those utilizing the Applequist model (such as AMBER ff02/ff02EP[12] and ff02r1[13]), the short-range 1−2 and 1−3 interactions are excluded to reduce the potential of "polarization catastrophe". Cieplak et al.[11] found that both f_e and f_t in eq 3 approach 1.0 when the distance between two interaction sites goes beyond 3.0 Å for the aforementioned four dipole interaction models.

In this work, a set of induced dipole models has been developed to reproduce the high-quality experimental molecular polarizabilities. We classified these models into five categories based on the different treatment of 1−2 (bonded), 1−3 (separated by two bonds), and 1−4 (separated by three bonds) interactions. In sets A, B, and C, the short-range 1−2 and 1−3 interactions are excluded, whereas they are included in sets D and E. Specifically, in set A, both 1−2 and 1−3 interactions are excluded, and the 1−4 interactions are included without scaling (i.e., 100%). In set B, the 1−2 and 1−3 interactions are excluded, and the 1−4 interactions are scaled down by 1.2. In set C, 1−2 and 1−3 interactions are excluded, and the 1−4 interactions are scaled down by 2. All 1−2, 1−3, and 1−4 interactions are fully considered in D and E sets. The difference between D and E sets is that they were obtained from different data sets.

Dipole Interaction Model Development. *Data Sources.* In experiment, the average polarizability of a molecule can be obtained by measuring dielectric constant and refractive index. The available atomic polarizability parameter sets in molecular mechanical force fields have been obtained by fitting to data sets that are quite small (fewer than 100 molecules).[15] The diversity of the training set molecules is also a concern. Recently, Bosque and Sales reported a 420-molecule data set with coverage on diverse functional groups in organic chemistry, including hydrocarbons (aliphatic and aromatic, cyclic and acyclic), alcohols, phenols, ethers, esters, peroxides, aldehydes, ketones, carboxylic acids, amines, imines, amides, nitriles, nitro-derivatives, disulfides, thiophenes, sulfides, sulfoxides, sulfones, phosphates, and halides.[20] Six molecules were removed because of duplication or apparent errors. The experimental polarizabilities α of the 420 molecules were obtained by measuring refractive indexes R, which relate to α by the Lorentz−Lorenz equation (eq 1). The experimental refractive indexes R were measured at 20 or 25 °C at the D-line of sodium wavelength (5893 Å). The entire data set was randomly divided into a training set (entries 1−335 in the Table S1 of the Supporting Information) and a test set (entries 336−420 in Table S1 of the Supporting Information) for model validation purpose. The compound names and the SMILES notation of the 420 molecules are listed in the Table S1 of the Supporting Information, and the experimental molecular polarizabilities are listed in Table S2 of the Supporting Information. Besides the Bosque−Sales data set, we also investigated the 70-molecule data set used by van Duijnen and Swart in their development of dipole interaction modes with the Thole's schemes.[15] The compound names and the SMILES notation of the 70 molecules are listed in Table S3 of the Supporting Information, and the experimental molecular polarizabilities are listed in Table S4 of the Supporting Information. This data set was further divided into a 52-molecule training set and an 18-molecule test set. Another subset, the 34-molecule validation set, which excludes small molecules that have only 1−2 and 1−3 interactions, was constructed to evaluate models trained using the Bosque and Sales data set. It is noted that the molecular polarizabilities in this study were converted to $Å^3$ unit for convenience of the later MM calculations. The converting factor from atomic unit (au) to $Å^3$ is 0.1481847114 using the Bohr radius of 0.5291772108 Å.

Optimization Protocol. GA is an efficient stochastic optimization method that has been widely applied to solve minimization problems such as conformational searches, molecular docking, as well as QSAR or QSPR analysis.[25−30] GA maximizes a fitness function through three operations that mimic natural evolution and selection, namely, mutation, crossover, and selection. First of all, a set of "chromosomes", which might be possible solutions to a question, are randomly generated. The "genes" in a "chromosome" correspond to the variables in question. For each "chromosome", the fitness is evaluated by a fitness function. The higher the score, the better the fitness, and the closer to the real answer it is. New "chromosomes" are then generated through crossover and mutation operations. In the subsequent selection operation, "chromosomes" with high fitness are evolved to the next generation, and those having low fitness are allowed to perish. The three operations are iteratively performed until a termination criterion is met. The power of GA lies in its ability to efficiently deal with multiple dimension or nonlinear problems.

The following are the important parameter values that control the GA optimization: (1) population size: the number of chromosomes in one generation −30; (2) chromosome size: the number of variables in question, specified in the last column of Table 1; (3) elite size: the number of elite "chromosomes", which enter the next generation directly −2; (4) random size: the number of "chromosomes" that have lowest fitting scores that are regenerated before selection −1; (5) mutation probability: the probability of performing mutation on each gene of each chromosome −0.01; (6) crossover probability: the probability of performing crossover on each "chromosome" in a population −0.90; (7) selection methods: tournament is used in the optimization stage;[31] (8) tournament number: number of "chromosomes" participating in selection each time −3; and (9) maximum iteration: the maximum iteration of optimization −10 000. For each parameter, the value followed by a dash symbol was

Table 1. Summary of the Polarizability Models[a]

no.	model[b]	data set	screening function	short-range interaction scaling factors			No. of parameters[g]
				1−2	1−3	1−4	
1	AL	BS[e]	linear	0.0	0.0	1.0	15 + 1
2	AE	BS	exponential	0.0	0.0	1.0	15 + 1
3	AT	BS	Tinker-like	0.0	0.0	1.0	15 + 1
4	AA	BS	Applequist	0.0	0.0	1.0	15 + 0
5	BL	BS	linear	0.0	0.0	1/1.2	15 + 1
6	BE	BS	exponential	0.0	0.0	1/1.2	15 + 1
7	BT	BS	Tinker-like	0.0	0.0	1/1.2	15 + 1
8	BA	BS	Applequist	0.0	0.0	1/1.2	15 + 0
9	CL	BS	linear	0.0	0.0	0.5	15 + 1
10	CE	BS	exponential	0.0	0.0	0.5	15 + 1
11	CT	BS	Tinker-like	0.0	0.0	0.5	15 + 1
12	CA	BS	Applequist	0.0	0.0	0.5	15 + 0
13	DL	BS	linear	1.0	1.0	1.0	15 + 1
14	DE	BS	exponential	1.0	1.0	1.0	15 + 1
15	DE_valid[c]	BS	exponential	1.0	1.0	1.0	15 + 1
16	DT	BS	Tinker-like	1.0	1.0	1.0	15 + 1
17	DA	BS	Applequist	1.0	1.0	1.0	15 + 0
18	DLE	BS	linear	1.0	1.0	1.0	10 + 1
19	DEE	BS	exponential	1.0	1.0	1.0	10 + 1
20	DAE	BS	Applequist	1.0	1.0	1.0	10 + 0
21	EL	vDS[f]	linear	1.0	1.0	1.0	11 + 1
22	EE	vDS	exponential	1.0	1.0	1.0	11 + 1
23	EA	vDS	Applequist	1.0	1.0	1.0	11 + 0
24	ELE	vDS	linear	1.0	1.0	1.0	9 + 1
25	ELE_vDS[d]	vDS	linear	1.0	1.0	1.0	9 + 1
26	EEE	vDS	exponential	1.0	1.0	1.0	9 + 1
27	EEE_vDS[d]	vDS	exponential	1.0	1.0	1.0	9 + 1
28	EAE	vDS	Applequist	1.0	1.0	1.0	9 + 0

[a] All polarizability models are divided into groups depending on the way the short-range interactions are treated. For A, B, and C sets, the 1−2 and 1−3 interactions were excluded. In set A, the 1−4 interactions were not scaled (100%). In set B, the 1−4 interactions were scaled down by 1.2. In set C, the 1−4 interactions were scaled down by 2.0. In sets D and E, the 1−2, 1−3, and 1−4 interactions were taken into account. [b] For each model name, the first letter designates the interaction sets, as explained above. The second letter denotes the screening function types: A, Applequist; L, Thole linear; E, Thole exponential; and T, the exponential screening function used in the Tinker force field. By default, a model was developed using an atom type scheme. However, when the third letter of the model name is 'E', each element has only one atom type. [c] DE_valid: means the model is developed with a training set for testing purpose. [d] ELE_vDS and EEE_vDS are models developed by van Duijnen and Swart. [e] BS: Bosque−Sales data set. [f] vDS: van Duijnen−Swart data set. [g] Integers before "+" are the number of atom types, and the integers after "+" are the number of screening factors.

applied in this work. As a flexible stochastic optimization algorithm, the performance of GA is controlled by many parameters. The most important parameters include the population size and the probabilities of performing mutation, crossover, and so on. The parameter set used in this study is suitable to study a variety of optimization problems[25−27] and is well-balanced in the computational cost and optimization efficiency. To assess the above parameter set and to find out if GA can reach or come close to the global minima, for each polarizable model, we performed GA optimizations at least six times, and the best parameter set was adopted.

A real number-encoded GA developed by ourselves was extensively used to optimize the atomic polarizabilities α_i as well as the screening factor, a, for four dipole interaction models, that is, the Applequist's, the Thole's linear, the Thole's exponential, and the Thole's Tinker-like exponential models. The fitness function of the GA optimization is defined in eq 10, where X, the average percent error, is calculated with eq 11. We added 0.01 to avoid possible floating-point error when X approaches 0.

$$\text{fitness} = \frac{1}{(0.01 + X)} \tag{10}$$

$$X = 100 \sum_{i=1}^{N} (|\alpha_i^{\text{calcd}} - \alpha_i^{\text{exptl}}|/\alpha_i^{\text{exptl}})/N \tag{11}$$

where α_i is the polarizability of the ith molecule and the superscripts "calcd" and "exptl" denote the calculated and experimental values, respectively. N is the number of molecules investigated. X is the average percentage error.

To construct reliable models and to avoid overfitting, we initially assigned each element (H, C, N, O, F, Cl, Br, I, S, P) to only one atom type. Additional atom types were added only if they demonstrated significant reduction in the errors.

In total, we have developed 26 dipole interaction models that belong to five model categories. Two published models by van Duijnen and Swart were used to calibrate our computer code.[15] The detailed information on those models is listed in Table 1. For a model name, the first letter designates its interaction type, and the second letter denotes the screening function type: A, Applequist; L, Thole linear; E, Thole exponential; and T, the exponential screening function used in the Tinker force field. By default, a model was developed using an atom type scheme. However, when the third letter of the model name is "E", each element shares only one atom type for this model. Finally, "vDS" means the model is developed by van Duijnen and Swart.

Model Cross Validations. First of all, using Model DE as an example, we developed a model with 335 training set molecules (entries 1−335 in Table S1 of the Supporting Information) of the Bosque−Sales' data set. Then, it was used to make predictions of molecular polarizabilities for the 85 molecules in the test set (entries 336−420).

Second, to investigate further the reliability of polarizability models developed using the Bosque−Sales' data set (sets A, B, C, and D), molecular polarizability prediction was conducted for two subsets of the van Duijnen−Swart data set, the 18-molecule test set, and the 34-molecule validation set.

4. RESULTS AND DISCUSSION

In this Article, we have developed a set of dipole interaction models that employ different screening functions and utilize different schemes of treating the 1−2, 1−3, and 1−4 interactions. We hoped that among these models, one or two outperform the others in the subsequent evaluation. The top performed models will be used to develop polarizable molecular mechanical force fields. The systematic evaluation of the polarizable models in molecular mechanics is presented in the concurrent article.

GA has been applied to optimize the atomic polarizability and the screening factor parameters. As a stochastic optimization method, GA may not reach the global minima in every run. Therefore, six GA runs were performed for each model. Encouragingly, for all models, all or most runs achieved similar results both for the atomic polarizability and screening factor parameters as well as for the fitness scores, indicating that our GA protocol was adequate and it reached or was close to the global minima in most cases. For example, for model AL, the APEs for six GA runs are 1.26, 1.23, 1.26, 1.24, 1.21, and 1.22%, respectively.

A, B, and C Model Sets (1−2 and 1−3 Interactions Excluded). One popular approach in molecular mechanical force fields is to exclude the 1−2 and 1−3 interactions from electrostatic calculations because their contribution to the polarization can be effectively absorbed by the long-range (1−4, 1−5, and so on) polarization through proper atomic polarizability and charge parametrization. Thus, in sets A, B and C models, the short-ranged 1−2 and 1−3 interactions are excluded. Moreover, the 1−4 interactions are scaled down in many force fields, including AMBER ff94/ff99.[32,33] The 1−4 interactions in sets B and C models are also scaled down by 1.2 and 2.0, respectively, and retained 100% in set A. It is noted that in sets A, B, and C models, oxygen has two atom types: sp^3 oxygen in hydroxyl, ether, and ester and sp^2 oxygen in the carbonyl group. Given the fact that sets A, B and C models treat the 1−2, 1−3, and 1−4 interactions in a similar fashion as the molecular mechanical force fields, these models are suitable for developing polarizable force fields.

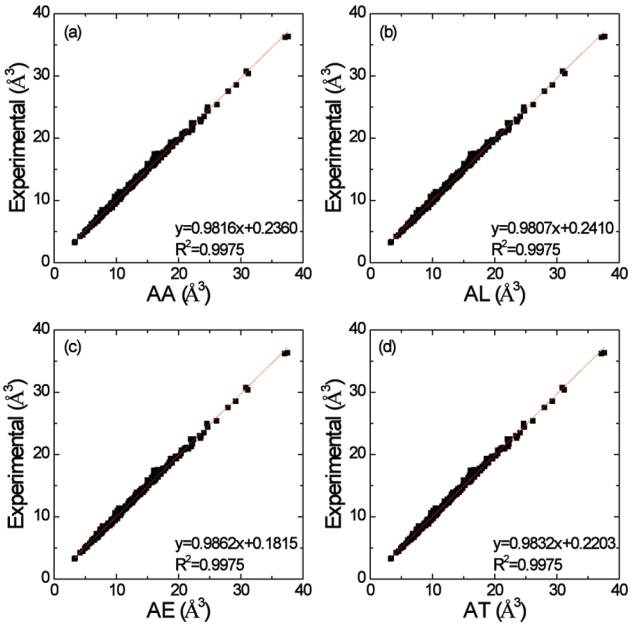


Figure 1. Scatter plots of calculated versus experimental polarizability (in Å^3) of the 420-molecule Bosque-Sales data set with the corresponding regression equation: (a) model AA, (b) model AL, (c) model AE, and (d) model AT.

When 1−2 and 1−3 interactions are omitted, although the Thole models outperform the Applequist model in individual model sets (sets A, B and C), the performance of all models in the three model sets is similar to each other for all four types of screening functions. The APEs are in the range from 1.15 to 1.23%, and the RMSEs are in the range from 0.225 to 0.250 Å^3. This is understandable because all four types of screening functions have similar f_e and f_t values in eqs 4−6 when the distance between two interaction sites goes beyond 3 Å. When the performance of the same screening function in different model sets is compared, a general trend is that the set C model performs slightly better than the set B model, and the latter is slightly better than the set A model. This performance trend is reasonable because more 1−4 interactions are scaled down in set C (scaling factor is 2) than set B (scaling factor is 1.2) and 100% 1−4 interactions are retained in set A.

The behavior of the models in the reduced interaction model sets (sets A, B, and C) is very different from the full interaction models in which different screening functions perform quite differently and the three Thole models apparently outperform the Applequist models. The plots of the experimental versus calculated molecular polarizability for the four set A models are shown in Figure 1. The fitting results of sets A, B, and C polarizability models are summarized in Table 2, and the atomic polarizability parameters and the screening factors are listed in Table 3. As shown in Table 3, the atomic polarizability parameters for the Applequist and three Thole models are essentially similar for the same model set. As for the same model in different model sets, the variation in the atomic polarizabilities is consistent with the scaling factors of the 1−4 interactions. Taking the Thole linear model as an example, for each of the major atom types (C2, C3, H, O2, O3, N), the atomic polarizability of BL is always between those of AL and CL.

Set D Models (1−2, 1−3, and 1−4 Interactions Included). Taking the advantage of a 420-molecule data set reported by

Table 2. Summary of the Fitting Results for A, B, and C Polarizability Models

model set	set A				set B				set C			
model	AL	AE	AT	AA	BL	BE	BT	BA	CL	CE	CT	CA
AUE (Å^3)	0.156	0.154	0.156	0.158	0.151	0.150	0.151	0.155	0.143	0.143	0.145	0.146
RMSE(Å^3)	0.250	0.243	0.247	0.250	0.240	0.237	0.237	0.243	0.226	0.225	0.228	0.229
APE (%)	1.21	1.21	1.22	1.23	1.19	1.19	1.19	1.22	1.15	1.16	1.16	1.17

Table 3. List of Atomic Polarizability Parameters and the Screening Factors for A, B, and C Models

model	AL	AE	AT	AA	BL	BE	BT	BA	CL	CE	CT	CA
C1[a]	1.3916	1.3759	1.3714	1.3998	1.3717	1.3814	1.3828	1.4260	1.4037	1.4049	1.3948	1.4296
C2[b]	1.2955	1.3297	1.2972	1.2808	1.3093	1.3280	1.3030	1.2758	1.3260	1.3226	1.3315	1.2929
C3[c]	0.9399	0.9937	0.9325	0.9158	0.9545	0.9888	0.9483	0.8896	0.9894	0.9738	1.0048	0.9136
H	0.4255	0.4087	0.4273	0.4356	0.4215	0.4120	0.4231	0.4516	0.4116	0.4199	0.4054	0.4466
NO[d]	1.4824	1.5385	1.5034	1.4574	1.5119	1.5372	1.5006	1.4626	1.5306	1.5118	1.5210	1.4795
N[e]	0.9603	0.9984	0.9727	0.9505	0.9677	0.9955	0.9745	0.9285	0.9788	0.9762	0.9851	0.9378
O2[f]	0.6049	0.5852	0.5993	0.6157	0.5936	0.5831	0.6016	0.6061	0.5856	0.5924	0.5906	0.5984
O3[g]	0.6148	0.6331	0.6273	0.6085	0.6222	0.6277	0.6283	0.6005	0.6267	0.6213	0.6280	0.6017
F	0.4839	0.4646	0.4856	0.4936	0.4791	0.4676	0.4796	0.5100	0.4655	0.4758	0.4572	0.4996
Cl	2.3707	2.3675	2.3664	2.3613	2.3730	2.3659	2.3641	2.3698	2.3667	2.3635	2.3628	2.3756
Br	3.5016	3.4894	3.5037	3.4976	3.4977	3.4898	3.4997	3.5056	3.4925	3.4930	3.4873	3.5040
I	5.5788	5.5747	5.5864	5.5746	5.5803	5.5736	5.5842	5.5684	5.5739	5.5732	5.5717	5.5767
S4[h]	2.3149	2.3409	2.3477	2.3097	2.5193	2.3601	2.3392	2.3323	2.3707	2.3649	2.3652	2.3409
S[i]	3.1686	3.1934	3.1780	3.1613	3.1740	3.1881	3.1774	3.1490	3.1903	3.1915	3.1872	3.1782
P	1.7927	1.8128	1.7894	1.7885	1.8204	1.8472	1.8066	1.7941	1.8556	1.8473	1.8583	1.8268
screening factor	2.5874	0.4766	1.6209		2.5939	0.4543	1.5171		2.5795	0.3744	1.5640	

[a] C1: sp^1 carbon. [b] C2: sp^2 carbon. [c] C3: sp^3 carbon. [d] NO: nitro nitrogen. [e] N: non-nitro nitrogen. [f] O2: sp^2 oxygen. [g] O3: sp^3 oxygen. [h] S4: S in sulfone. [i] S: nonsulfone S.

Bosque and Sales, we developed eight dipole interaction models. In DLE, DEE, and DAE models, each of the 10 elements (H, C, N, O, P, S, F, Cl, Br, I) has one atom type. The other five models, DL, DE, DT, DA, and DE_valid, employ 15 atom types, including three for carbon (sp^1, sp^2, and sp^3), two for nitrogen (nitrogen in nitro group and the other nitrogen atoms), two for sulfur (sulfur in a sulfone group and the other sulfurs), two for oxygen (sp^2 oxygen and sp^3 oxygen), and a single atom type for the other elements.

Starting from the models based on chemical composition (DLE, DEE, and DAE), sp^1, sp^2, and sp^3 were first differentiated. This led to three intermediate models for all three dipolar interaction models. The APE significantly dropped from 6.83 to 4.77% for the Applequist's, from 2.62 to 1.83% for the Thole's linear, and from 2.73 to 1.62% for Thole's exponential models, respectively. We then examined the error sources of the intermediate models and found that many nitro-derivatives and sulfur-containing compounds had large prediction errors. Therefore, we decided to separate nitrogen in nitro groups from the other nitrogen atoms and to separate sulfur in sulfone functional groups from the other sulfur type as well. Finally, four new dipole interaction models that utilize the above atom type definition scheme were developed, which are DL, DE, DT, and DA.

The fitting errors of these models are listed in Table 4. It is concluded that the models with 14 atom types outperform those models that are based on chemical composition (i.e., each element shares one atomic polarizability parameter). The best model, DE has AUE, RMSE and APE of 0.166 Å^3, 0.250 Å^3 and 1.30%, respectively. In general, all three Thole's models achieve much better performance than the Applequist's models suggesting the critical roles that the screening function plays in the Thole models when short-range polarization is included. The atomic polarizability parameters, α_i, and the screening factor parameter, a, are listed in Table S. Unlike sets A, B and C models, there is much larger variation in the atomic polarizabilities as a function of model type for this model set.

Set E Models (1−2, 1−3, and 1−4 Interactions Included). The set E models are similar to set D models, except set E models were developed using the van Duijnen−Swart data set[15] and the set D models were developed using the Bosque−Sales data set.[20] Following the approach of van Duijnen and Swart,[15] the 70-molecule data set is divided into training and testing sets. The 52-molecule training set was used to construct six models using a GA to optimize a fitness function described by eq 10. Three models, that is, EAE, ELE, and EEE, each utilized nine atom types (each element has one atom type) to fit the molecular polarizability calculated by a dipole interaction model to the experimental values. The APEs for the Applequist's, the Thole's linear, and the Thole's exponential models are 14.78, 3.31, and 3.58%, respectively. The calculated molecular polarizabilites using the two published models (ELE_vDS and EEE_vDS) are listed in Table S4 of the Supporting Information for comparison. The molecular polarizabilites calculated with our program are close to the numbers reported by van Duijnen and Swart. The negligible difference was caused by slightly different geometries. The APE values for ELE_vDS and EEE_vDS are 3.68 and 3.70, respectively.

dx.doi.org/10.1021/jp112133g |J. Phys. Chem. B 2011, 115, 3091–3099

Table 4. Fitting Results of D Models

model	DL	DE	DA	DT	DLE	DEE	DAE
AUE (Å^3)	0.194	0.166	0.446	0.196	0.345	0.362	0.726
RMSE (Å^3)	0.280	0.250	0.674	0.285	0.550	0.539	1.528
APE (%)	1.55	1.30	3.82	1.58	2.62	2.73	6.83

Table 5. Atomic Polarizabilities and the Screening Factors for the D Models

atom type	DL	DE	DT	DA	DEL	DEE	DEA
C1[a]	1.6781	1.3668	1.7821	0.4193	2.2220	1.6677	0.5959
C2[b]	1.7032	1.5641	1.7838	0.6390	2.2220	1.6677	0.5959
C3[c]	1.1474	1.2223	0.9462	0.7070	2.2220	1.6677	0.5959
H	0.5881	0.4803	0.6617	0.2213	0.1922	0.2919	0.2833
NO[d]	2.2158	1.9506	2.4512	0.7517	1.8624	1.3928	0.3747
N[e]	1.2687	1.1609	1.2322	0.5011	1.8624	1.3928	0.3747
O2[f]	0.6418	0.5647	0.5094	0.3354	0.6657	0.5987	0.3512
O3[g]	0.6817	0.6339	0.6182	0.3849	0.6657	0.5987	0.3512
F	0.6089	0.4779	0.6095	0.4543	0.2413	0.3344	0.5234
Cl	2.4870	2.5034	2.4986	2.0420	2.4767	2.4642	2.1198
Br	3.5661	3.6422	3.5197	3.1131	3.7206	3.6401	3.1942
I	5.5408	5.7306	5.5096	4.8942	5.9896	5.7591	5.0242
S4[h]	3.0581	2.8744	2.9779	1.6789	4.0083	3.4604	1.7223
S[i]	3.6290	3.5159	3.5216	2.1900	4.0083	3.4604	1.7223
P	2.2689	1.9844	2.1370	0.9271	1.9275	1.9303	0.8777
screening factor	2.0580	0.4968	1.3774		2.1714	0.5132	

[a] C1: sp^1 carbon. [b] C2: sp^2 carbon. [c] C3: sp^3 carbon. [d] NO: nitro nitrogen. [e] N: non-nitro nitrogen. [f] O2: sp^2 oxygen. [g] O3: sp^3 oxygen. [h] S4: S in sulfone. [i] S: nonsulfone S.

Through examination of the prediction errors of those models utilizing elements as atom types (EAE, ELE, EEE, ELE_vDS, and EEE_vDS), we found that the calculated polarizabilities of alkenes and benzene derivatives were systematically lower than the experimental values. We thus introduced three carbon atom types: sp^1, sp^2, and sp^3 in the fitting procedure, leading to three other models, that is, EA, EL, and EE. Each of the three models has eleven atom types. The performance of the three models was considerably improved, and the APE values were 6.86, 2.93, and 3.42% for the Applequist's, the Thole's linear, and the Thole's exponential models, respectively. The calculated molecular polarizabilities of the eight set E models are listed in Table S4 of the Supporting Information, and the fitting results are summarized in Table 6. The atomic polarizability parameters, α_i, and the screening factor, a, are listed in Table 7. Similar to set D, there is a larger variation in the atomic polarizabilities as a function of model type for model set E.

The eight models were assessed against the 18 molecules in the test set, and the results are listed in Table S4 of the Supporting Information. The overall performance of the two Thole's models, ELE and EEE, is comparable to the two models reported by van Duijnen and Swart, whereas the other two Thole's models EL and EE that divide the carbons into sp^1, sp^2, and sp^3 atom types perform much better in terms of AUE, RMSE, and APE. Similarly, the EA outperforms EAE significantly.

Cross validation is a necessary step to investigate the reliability of the models. In this work, two types of validations were conducted. First, the model DE was reoptimized by GA using only 335 molecules of the Bosque−Sales data set as the training set. For the training set molecules (entries 1−335 in Table S1 of the Supporting Information), the AUE, RMSE, and APE are 0.169 Å^3, 0.255 Å^3, and 1.35%, respectively. Then, this model was applied to predict molecular polarizabilites of the 85 molecules in the test set (entries 336−420 in Table S1of the Supporting Information), the AUE, RMSE, and APE are 0.196 Å^3, 0.292 Å^3, and 1.35%, respectively. The results are listed in Table S1 of the Supporting Information under the column "DE_valid". The results show that DE is a reliable model because the APE of the test set molecules is identical to that of the training set molecules (1.35%), giving us the confidence that our models were free from deficiency due to potential overfitting.

Table 6. Fitting Results of E Models Using the 52-Molecule van Duijnen−Swart's Training Data Set

model	EL	EE	EA	ELE	ELE_vDS	EEE	EEE_vDS	EAE
AUE (Å^3)	0.167	0.215	0.430	0.221	0.240	0.240	0.236	0.596
RMSE (Å^3)	0.384	0.465	0.950	0.489	0.493	0.532	0.507	0.852
APE (%)	2.93	3.42	6.86	3.31	3.68	3.58	3.70	14.78

The 16 models in sets A, B, C, and D, developed using the Bosque−Sales' data set, were further evaluated by comparing with the van Duijnen−Swart data set. There are total of 70 molecules in the van Duijnen−Swart data set, divided into two sets: the 18-molecule test set and the 52-molecule main set. It should be emphasized that neither data set is used in the fitting of A−D model sets. Therefore, both can be used to validate the models. However, among the 70 molecules in the van Duijnen−Swart data set, 36 are small and simple molecules that contain only 1−2 and 1−3 interactions and are unsuitable to test A−C model sets because the 1−2 and 1−3 interactions are all excluded in these models. Therefore, these molecules are excluded in the tests. The remaining 34 molecules are referred to as the validation set and are used to validate the models. Because the 18-molecule test set was used to validate the set E models, for comparison, we also calculated the polarizabilities of these molecules, even though 8 of the 18 are simple molecules that have only 1−2 and 1−3 interactions, and the results are listed under the title of "18-molecule test set".

The calculated molecular polarizabilites are listed in Table S4 of the Supporting Information. Comparison between the predicted and experimental polarizabilites is summarized in Table 8. For the 18-molecule van Duijnen−Swart test set, it is encouraging that all sixteen models except model DA achieved notably better results than any of the eight set E models in terms of AUE, RMSE, and APE. Therefore, although the A−C set models do not include the 1−2 and 1−3 interactions, they can be used to represent the simple molecules that have only 1−2 and 1−3 interactions. For those simple molecules like water, the off-diagonal elements of the dipole field tensor matrixes become zero as 1−2 and 1−3 interactions are turned off. The molecular polarizabilities are calculated by summing up the diagonal elements. The molecular polarizability of water is 1.48, 1.47, 1.45, and 1.48 Å^3 for AA, AL, AE, and AT, respectively. The predicted values are very close to the experimental value, 1.49 Å^3. The set B and C models perform similarly in predicting the polarizability of water.

As for the 34-molecule validation set, most models apparently outperform the best set E model, EL. This is remarkable because these molecules were not used in the model-fitting process. It shows that our polarizability models are highly transferable. All 12 models in sets A, B, and C except BA have better APE than EL (2.25% for BA versus 2.22% for EL). For the four D models, DE outperforms EL, and the performance of DL and DT is slightly worse than that of EL in term of APE. A similar pattern was found for AUE and RMSE as well. In conclusion, although the models in sets A, B, C, and D were developed using the Bosque−Sales data set, tests on the independent van Duijnen−Swart data set indicates that

Table 7. Atomic Polarizabilities and the Screening Factors of the Set E Models Based on the van Duijnen−Swart Data Set

atom types	EL	EE	EA	ELE	ELE_vDS	EEE	EEE_vDS	EAE
C1 (sp^1)	1.4722	1.2565	0.3861	1.5292	1.5079	1.3178	1.2886	0.5566
C2 (sp^2)	1.6946	1.3567	0.6366	1.5292	1.5079	1.3178	1.2886	0.5566
C3 (sp^3)	1.5030	1.2165	0.7756	1.5292	1.5079	1.3178	1.2886	0.5566
H	0.4960	0.4550	0.1686	0.5190	0.5189	0.4402	0.4138	0.2720
F	0.5196	0.4612	0.3613	0.4854	0.4359	0.4464	0.4447	0.2738
Cl	2.4767	2.3901	2.0060	2.4306	2.3879	2.4000	2.4003	1.9835
Br	3.5338	3.4658	2.9634	3.4389	3.3589	3.4736	3.4929	2.9251
I	5.5109	5.4598	4.9106	5.3648	5.0743	5.4810	5.4811	4.9019
N	1.0746	0.9969	0.4994	1.1061	1.1269	0.9815	0.9716	0.3993
O	0.9377	0.7892	0.4662	0.9520	0.9475	0.8234	0.8520	0.3412
S	2.8694	2.4925	1.4164	2.8379	2.9255	2.4910	2.4744	1.8186
screening factor	2.0158	0.4761		1.8255	1.7278	0.5018	0.4694	

Table 8. Validation Using van Duijnen−Swart's Data Sets

model	18-molecule test set[a]			34-molecule validation set[b]		
	AUE (Å3)	RMSE (Å3)	APE (%)	AUE (Å3)	RMSE (Å3)	APE (%)
AA	0.191	0.258	2.84	0.164	0.236	2.19
AL	0.188	0.259	2.73	0.161	0.237	2.14
AE	0.185	0.248	2.56	0.158	0.235	2.12
AT	0.183	0.247	2.62	0.159	0.229	2.13
BA	0.197	0.262	3.14	0.168	0.236	2.25
BL	0.180	0.245	2.61	0.154	0.223	2.06
BE	0.183	0.242	2.60	0.157	0.233	2.13
BT	0.179	0.239	2.61	0.155	0.226	2.11
CA	0.189	0.248	3.07	0.162	0.231	2.23
CL	0.184	0.235	2.72	0.154	0.228	2.11
CE	0.180	0.234	2.73	0.155	0.229	2.13
CT	0.183	0.235	2.62	0.153	0.228	2.08
DA	0.294	0.383	5.55	0.516	0.783	8.98
DL	0.241	0.287	3.37	0.207	0.274	2.81
DE	0.203	0.255	2.77	0.165	0.238	2.16
DT	0.234	0.292	3.70	0.196	0.260	2.68
EA	0.806	1.277	14.83	0.630	0.978	6.79
EL	0.273	0.428	3.53	0.207	0.341	2.22
EE	0.317	0.390	3.70	0.236	0.371	2.56
ELE	0.366	0.565	4.16	0.347	0.574	3.51
ELE_vDS	0.375	0.596	4.13	0.351	0.589	3.53
EEE	0.403	0.529	4.39	0.319	0.543	3.23
EEE_vDS	0.347	0.449	3.82	0.300	0.497	3.14
EAE	1.514	2.39	24.28	0.929	1.276	9.72

[a] Original test set of van Duijnen−Swart. The set also includes eight molecules that have only 1−2 and 1−3 interactions. [b] 34 molecules selected from the 70 molecules of van Duijnen−Swart after excluding the 36 molecules that have only 1−2 and 1−3 interactions, including the 10 molecules listed in the 18-molecule test set.

the models are highly transferable and actually have smaller errors than the models developed using the van Duijnen−Swart data set. The improved performance of A−D model sets in comparison with set E models indicates that the increased number of parameters was justified and the models are highly transferable.

Furthermore, comparison among A, B, and C sets with D and E sets made it clear that the inclusion of the 1−2 and 1−3 terms notably reduced the accuracy of the model. In the A, B, and C sets, the RMSEs were all 0.25 $Å^3$ or less, AUEs were <0.16 $Å^3$, and APE's were no more than 1.23%. In comparison, in the D set models, the RMSEs ranged from 0.25 (DE) to 1.528 $Å^3$ (DAE), AUE ranged from 0.166 (DE) to 0.726 $Å^3$ (DAE), and APE was from 1.3 (DE) to 6.83% (DAE), all greater than those of the A, B, and C sets. The E set models were much worse than A, B, and C sets, with RMSEs ranging from 0.38 (EL) to 0.95 $Å^3$ (EA), AUEs from 0.167 (EL) to 0.596 $Å^3$ (EAE), and APE from 2.93 (EL) to 14.78% (EAE). This strongly suggests that inclusion of the 1−2 and 1−3 polarization significantly reduced the accuracy of the models. These observations are also consistent with the transferability of the A, B, and C model sets.

5. CONCLUSIONS

In this work, a set of dipole interaction models have been constructed using a data set recently published by Bosque and Sales. This 420-molecule data set covers most of the important chemical functional groups. Four types of screening functions, namely, the Applequist, the Thole linear, the Thole exponential, and the Thole Tinker-like exponential have been thoroughly

investigated. The models are grouped into five categories based on how 1−2, 1−3, and 1−4 interactions were treated. For the three reduced interaction model sets, sets A, B, and C, the performance of polarizability calculations is very encouraging: the average percent errors are in the range from 1.15 to 1.23%, and the root-mean-square errors are in the range from 0.225 to 0.250 Å^3. All models have been developed using an in-house GA, and all models are very reliable based on the results of cross validations and the consistent results of multiple GA runs. The critical assessment of these dipole interaction models will be presented in the concurrent article.

■ ASSOCIATED CONTENT

Supporting Information. Compound names and the SMILES notation of the Bosque−Sales and the van Duijnen−Swart data sets. Experimental values of molecular polarizabilities. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: duan@ucdavis.edu. Fax: (530)-754-9658. Tel: (530)-754-5625.

■ ACKNOWLEDGMENT

We are grateful to acknowledge the research support from the NIH (R01GM79383, Y.D., P. I.) and the TeraGrid for the computational time (TG-CHE090098 J.W., P.I., and TG-CHE090135 P.C., P.I.).

■ REFERENCES

(1) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. C.; Debolt, S. E.; Ferguson, D. M.; Seibel, G. L.; Kollman, P. A. *Comput. Phys. Commun.* **1995**, *91*, 1.
(2) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(3) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D., J; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(4) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
(5) Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497.
(6) Dehez, F.; Angyan, J. G.; Gutierrez, I. S.; Luque, F. J.; Schulten, K.; Chipot, C. *J. Chem. Theory Comput.* **2007**, *3*, 1914.
(7) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.
(8) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.
(9) Tan, Y. H.; Luo, R. *J. Chem. Phys.* **2007**, *126*, 094103.
(10) Tan, Y. H.; Tan, C. H.; Wang, J.; Luo, R. *J. Phys. Chem. B* **2008**, *112*, 7675.
(11) Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M. *J. Phys.: Condens. Matter* **2009**, 21.
(12) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1048.
(13) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, *27*, 781.
(14) Applequist, J.; Carl, J. R.; Fung, K. K. *J. Am. Chem. Soc.* **1972**, *94*, 2952.
(15) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399.
(16) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.
(17) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
(18) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.
(19) Elking, D.; Darden, T.; Woods, R. J. *J. Comput. Chem.* **2007**, *28*, 1261.
(20) Bosque, R.; Sales, J. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1154.
(21) Stout, J. M.; Dykstra, C. E. *J. Am. Chem. Soc.* **1995**, *117*, 5127.
(22) Wang, J. M.; Xie, X. Q.; Hou, T. J.; Xu, X. J. *J. Phys. Chem. A* **2007**, *111*, 4443.
(23) de Vries, A. H.; van Duijnen, P. T.; Zijlstra, R. W. J.; Swart, M. *J. Electron Spectrosc. Relat. Phenom.* **1997**, *86*, 49.
(24) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8543.
(25) Wang, J. M.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1219.
(26) Wang, J. M.; Krudy, G.; Xie, X. Q.; Wu, C. D.; Holland, G. *J. Chem. Inf. Model.* **2006**, *46*, 2674.
(27) Hou, T. J.; Wang, J. M.; Chen, L. R.; Xu, X. J. *Protein Eng.* **1999**, *12*, 639.
(28) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
(29) Xiao, Y. L.; Williams, D. E. *J. Phys. Chem.* **1994**, *98*, 7191.
(30) Bowie, J. U.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4436.
(31) Yang, J. P.; Soh, C. K. *J. Comput. Civ. Eng.* **1997**, *11*, 195.
(32) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(33) 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.