Constructing *ab initio* force fields for molecular dynamics simulations

Yi-Ping Liu, Kyungsun Kim, B. J. Berne, Richard A. Friesner, and Steven W. Rick[a)]
Department of Chemistry and Center for Biomolecular Simulations, Columbia University, New York, New York 10027

We explore and discuss several important issues concerning the derivation of many-body force fields from *ab initio* quantum chemical data. In particular, we seek a general methodology for constructing *ab initio* force fields that are ''chemically accurate'' and are computationally efficient for large-scale molecular dynamics simulations. We investigate two approaches for modeling many-body interactions in extended molecular systems. The interactions are adjusted to reproduce the many-body energy in small molecular clusters. Subsequently, the potential parameters affecting only pair interactions are then varied to reproduce the *ab initio* binding energy of dimers. This simple procedure is demonstrated in the design of a new polarizable force field of water. In particular, this new model incorporates the usual many-body interactions due to electrostatic polarization and a type of nonelectrostatic many-body interactions exhibited in bifurcated hydrogen-bonded systems. The static and dynamical properties predicted by the new *ab initio* water potential are in good agreement with the successful empirical fluctuating-charge potential of Rick *et al.* and with experiment. The aforementioned ''cluster'' approach is compared with an alternative method, which regards many-body interactions as manifestations of the electrostatic polarization properties of individual molecules. The effort required to build *ab initio* databases for force field parametrization is substantially reduced in this alternative method since only the monomer properties are of interest. We found intriguing differences between these two approaches. Finally our results point to the importance of discriminating *ab initio* data for force field parametrization. This is essentially a consequence of the simple functional forms employed to model molecular interactions, and is inevitable for large-scale molecular dynamics simulations. © *1998 American Institute of Physics.* [S0021-9606(98)51412-2]

I. INTRODUCTION

Large-scale numerical simulations of complex molecules serve as a bridge between experiments and analytic theories. The value of the predictions based on simulation is limited by the accuracy of the potential energy functions. Consequently, much research effort has been devoted to the design of potential energy functions. In general, simple functional forms are used in the design of potential energy functions in order to reduce the computational cost entailed by large-scale simulations.[1] The parameters in a potential model are usually derived either empirically or from *ab initio* data at various molecular configurations. The most accurate potential functions to date are derived empirically since higher-order many-body effects are inherent in the experimental data and are thus *effectively* incorporated in the fitted parameters. This approach, however, lacks generality due to the following limitations. First and foremost, it requires empirical input that is often unavailable or unattainable. Second, a set of parameters derived from empirical data for a certain thermodynamic state is not generally appropriate for simulations under a different set of conditions and different molecular environment. As has been recognized, transferability

cannot be rigorously realized unless many-body intermolecular interactions are explicitly treated. These obstacles may appear irrelevant with the advent of the *direct dynamics* methodology, which replaces force fields with energy and forces acquired directly from *ab initio* quantum chemical calculations whenever they are needed during a dynamics calculation. While much progress has been made in *ab initio* molecular dynamics methods, and useful insights have been obtained thereby, applications of *ab initio* molecular dynamics methods to simulations of long-time dynamics of complex extended molecular systems remain a distant future objective. Be it rationalization of biochemical functions or the intelligent design of new materials, simple force fields continue to be the most practical representation of intermolecular interactions. The challenge is how to construct ''chemically accurate'' *ab initio* force fields.

A prerequisite for *ab initio* force fields is accurate *ab initio* input for parametrization. Quantum chemical methods for determining many-body intermolecular interactions include perturbation methods and the supermolecule approach. The former evaluates the intermolecular interactions directly as perturbations, whereas in the latter, the intermolecular interaction energy is calculated as the energy difference between the whole system and the sum of the constituent molecules in isolation.[2] Combined with energy decomposition analyses,[3,4] perturbation methods provide information on energy components, such as electrostatic, exchange, induction,

dispersion, etc., in addition to the total intermolecular interaction energy. Such information not only advances our understanding of the nature of intermolecular interactions, but has also been exploited in the design of functional forms for various components of intermolecular interactions. Several *ab initio* water potentials under development[5–7] may be roughly classified in this category. Prior energy decomposition analyses on several prototype systems have shown that each energy component is usually large and the interaction energy resulting from their superposition is small due to cancellation.[4,8] Thus, a prerequisite of this type of force-field fitting procedure is the availability of accurate estimates of each energy component. The estimates of these energy components are, however, less accurate than the interaction energy itself based on *ab initio* quantum chemical calculations.[9] These energy components exhibit varying complex dependence on nuclear coordinates and must be fitted with complicated functional forms. As such, while very accurate potential functions have been obtained for some small molecules, these functions must be substantially simplified for molecular dynamics simulations,[7] and thus additional ambiguity is introduced. Finally, the convergence in the perturbation expansion is always a serious consideration. The supermolecule approach, on the other hand, provide a direct estimate of intermolecular interaction energy. In addition, it is straightforward to include dynamic electron-correlation corrections with the localized molecular orbitals. Thus the supermolecule approach is well suited for providing *ab initio* databases for parametrizing force fields, particularly when the dominant component of the intermolecular interaction is known.

Perhaps the most challenging aspect in the quest for accurate force fields is *how* and to *what extent* complex reality may be reduced to accommodate the computational cost, and to achieve transferability without undue numerical expense for an individual system of interest. The simplified representation of molecular interactions resulted from such a reduction would determine the *ab initio* input needed for force field parametrization, and is inextricably linked to the strategy for acquiring an *ab initio* database as well as simulation techniques. The explicit inclusion of many-body interactions, for example, may be approached with two different mapping strategies. In the first method, it is assumed that the potential function parameters pertinent to many-body interactions in an extended molecular systems should also give a good representation of the many-body interactions in small molecular clusters. Thus, the many-body interactions in small molecular clusters are partitioned progressively into terms involving three bodies, four bodies, etc. If this sequence converges within a reasonably small number of terms, parametrization can be carried out for each term in the sequence up to the order of satisfactory convergence. This procedure is logical and conceptually simple. Substantial effort is, however, needed to build the *ab initio* database. In particular, the determination and parametrization of each term in the sequence becomes exceedingly cumbersome for large molecules. In the second procedure, many-body interactions are regarded largely as manifestations of the properties of individual molecules. Therefore, the *ab initio* input for potential-function

parametrization can simply be derived from calculations on monomers. According to prior energy decomposition analyses, electrostatic polarization is found to be the dominant many-body interaction, and several *ab initio* water potentials under development model many-body interactions by including molecular polarizability in the force fields.[6,7,10]

Finally, we note that the reduction in the representation of intermolecular interactions inevitably leads to the loss of one-to-one correspondence between the fitting parameters and the fitted properties. In modeling electrostation polarization, for example, the forces between molecules close to each other are not purely electrostatic, but a mixture of both the electrostatic and nonelectrostatic interactions. The solution to this problem is to construct a database free of inappropriate samples, or to perform a weighted parametrization. This strategy can be easily executed for the above example, but it is not well defined when applied to fitting other properties of the monomer, dimer, or clusters. Research addressing this issue is relatively rare. Yet, understanding this problem is of great importance to establishing a general procedure for deriving accurate general-purpose *ab initio* force fields.

This paper is our first step toward the goal of constructing accurate *ab initio* force fields. The immediate objective is to examine the aforementioned important issues, with an emphasis on intermolecular many-body interactions. Using water as an example, we calculate the properties of the water monomer, as well as dimers and trimers at various intermolecular configurations. The two approaches of parametrizing many-body interactions mentioned in the preceding paragraphs are compared, and their relation to each other is commented on. A major part of this work is focused on parametrizing a new *ab initio* water potential as a proof of concept, and we calculated the liquid-state properties of water, including the structure, diffusion constants, and frequency-dependent dielectric constants using molecular dynamics simulations. Inspired by the recent success of the empirical water potential of Rick, Stuart, and Berne,[11] we have elected to approach electrostatic polarization by the so-called *principle of electronegativity equalization*.[12] The empirical fluctuating charge potential of Rick *et al.* gives accurate gas-phase dipole moment and liquid-state properties, including radial distribution functions, dielectric constants, and the diffusion constant. Moreover, in the same spirit as *ab initio* molecular dynamics methods, the fluctuations of the charges in their model are treated as dynamical variables and propagated in time as the nuclear coordinates, resulting in an efficient molecular dynamics algorithm with computational cost (CPU time) only 1.1 times of that for the nonpolarizable water models.

We choose water as our prototype system because of its polarizability and hydrogen bonding capability. The cooperative aspect of hydrogen bonding in water was first suggested by Frank and Wen.[13] Research on water in the past two decades shows that this kind of many-body effect is responsible for the three-dimensional sequential hydrogen-bonded network in liquid water, and that the anomalous properties of water arise from the competition between two packing patterns, the relatively bulky local structures with nearly tetrahedral angles and strong bonds, and the more compact ar-

rangements exhibiting more strain and undergoing bond breaking.[14]

This paper is organized as follows. In Sec. II, we give a brief review of the electronegativity equalization principle; the fluctuating charge (FQ) model is reformulated to illuminate its properties, in particular, the polarization response of this model; and the empirical FQ potential of Rick *et al.* is briefly discussed. Computational details are given in Sec. III. In Sec. IV, we analyze the polarization response of the FQ model with respect to uniform and nonuniform external fields; in particular, we use the water molecule as an example and compare the FQ model to *ab initio* quantum chemical calculations. This is followed by an examination of the three-body interactions in water trimers obtained from quantum chemical calculations and the FQ model. We present our new *ab initio* FQ potential for water, including both the many-body and the pair components in Sec. V. In particular, we introduce a new functional form to model the short-range nonelectrostatic many-body effects, and we parametrize the many-body interactions to the three-body energies from quantum chemical calculations. The static and dynamic properties predicted using the new *ab initio* FQ potential are discussed in this section. Several observations from our study are discussed in Sec. VI, followed by our conclusions.

II. PROPERTIES OF DYNAMICAL FLUCTUATING CHARGE MODEL

In the FQ model, molecules respond to their environment by intramolecular charge transfer according to the *principle of electronegativity equalization*. This principle states that the instantaneous electronegativities of individual atoms that make up a molecule are equal.[12] The instantaneous electronegativity, χ_α, of an atom depends on the partial charge of the atom as well as the electrostatic environment surrounding the atom. Using a neutral atom as a reference point, the energy of an isolated atom can be expanded to the second order in charge as[15]

$$u(\{q_\alpha\}) = u_\alpha(0) + \tilde{\chi}_\alpha^0 q_\alpha + \frac{1}{2} J_{\alpha\alpha}^0 q_\alpha^2, \tag{1}$$

where the coefficients for the linear ($\tilde{\chi}_\alpha^0$) and the quadratic ($J_{\alpha\alpha}^0$) terms are characteristics of each particular atom type. The instantaneous electronegativity is defined as

$$\chi_\alpha = e \frac{\partial u}{\partial q_\alpha}. \tag{2}$$

Thus, the instantaneous electronegativity of an isolated neutral atom is equal to $\tilde{\chi}_\alpha^0$. If the partial derivative in Eq. (2) is approximated by a finite difference, $\tilde{\chi}_\alpha^0$ is found to be the ''Mulliken electronegativity,'' namely, one-half of the sum of the ionization potential and the electron affinity.[16] The value for $J_{\alpha\alpha}^0$ may be determined from the ionization potential and the electron affinity as well, and it is the so-called ''absolute hardness'' [17] of atom type α in isolation. In an aggregate of atoms, the total energy of the system consists of contribution from each atom and also interatomic interactions, and thus is a function of both the charges and the coordinates, $U(\{\mathbf{r}\},\{q\})$. Typically when there is no exter-

nally applied fields, the intra- and intermolecular potential energy $U(\{\mathbf{r}\},\{q\})$ for theoretical simulations is partitioned into Coulombic and nonelectrostatic types of interactions,

$$
\begin{aligned}
U(\{\mathbf{r}\},\{q\}) = & \sum_{i=1}^{N_{\text{mole}}} \sum_{\alpha=1}^{N_{\text{site}}} \left[\chi_{i\alpha}^0 q_{i\alpha} + \frac{1}{2} J_{\alpha\alpha}^0 q_{i\alpha}^2 \right] \\
& + \sum_{i=1}^{N_{\text{mole}}} \sum_{\alpha}^{N_{\text{site}}} \sum_{\beta>\alpha}^{N_{\text{site}}} J_{i\alpha i\beta} q_{i\alpha} q_{i\beta} \\
& + \sum_{i=1}^{N_{\text{mole}}} \sum_{j>i}^{N_{\text{mole}}} \sum_{\alpha}^{N_{\text{site}}} \sum_{\beta}^{N_{\text{site}}} J_{i\alpha j\beta} q_{i\alpha} q_{j\beta} \\
& + \sum_{i=1}^{N_{\text{mole}}} \sum_{j>i}^{N_{\text{mole}}} \sum_{\mu}^{N_{\text{atom}}} \sum_{\nu}^{N_{\text{atom}}} V(r_{i\mu j\nu}),
\end{aligned}
\tag{3}
$$

where there are N_{mole} molecules and each molecule consists of N_{atom} atoms and N_{site} charge sites; the terms $J_{i\alpha i\beta}$ and $J_{i\alpha j\beta}$ are related to the intra- and intermolecular electrostatic interactions, respectively; subscripts i and j denote the molecules, α and β denote the charge sites, and μ and ν denote the atoms. For the convenience of a later discussion, we denote the sum of the right-hand side of Eq. (3) without the $V(r_{i\mu j\nu})$ terms by U_{ee},

$$
\begin{aligned}
U_{ee}(\{\mathbf{r}\},\{q\}) = & \sum_{i=1}^{N_{\text{mole}}} \sum_{\alpha=1}^{N_{\text{site}}} \left[\chi_{i\alpha}^0 q_{i\alpha} + \frac{1}{2} J_{\alpha\alpha}^0 q_{i\alpha}^2 \right] \\
& + \sum_{i=1}^{N_{\text{mole}}} \sum_{\alpha}^{N_{\text{site}}} \sum_{\beta>\alpha}^{N_{\text{site}}} J_{i\alpha i\beta} q_{i\alpha} q_{i\beta} \\
& + \sum_{i=1}^{N_{\text{mole}}} \sum_{j>i}^{N_{\text{mole}}} \sum_{\alpha}^{N_{\text{site}}} \sum_{\beta}^{N_{\text{site}}} J_{i\alpha j\beta} q_{i\alpha} q_{j\beta}.
\end{aligned}
\tag{4}
$$

At each molecular configuration $\{\mathbf{r}\}$, the charges are redistributed among the atoms so as to equalize the instantaneous electronegativities. That is, the ground-state charge distribution $\{q^{\text{eq}}(\{\mathbf{r}\})\}$ satisfies

$$\left(\frac{\partial U}{\partial q_{i\alpha}} \right)_{\{q\}=\{q^{\text{eq}}\}} = \left(\frac{\partial U}{\partial q_{i\beta}} \right)_{\{q\}=\{q^{\text{eq}}\}}, \tag{5}$$

or equivalently,

$$\left(\frac{\partial U_{ee}}{\partial q_{i\alpha}} \right)_{\{q\}=\{q^{\text{eq}}\}} = \left(\frac{\partial U_{ee}}{\partial q_{i\beta}} \right)_{\{q\}=\{q^{\text{eq}}\}}, \tag{6}$$

and the ground-state energy $U_g(\{\mathbf{r}\})$ is equal to $U(\{q^{\text{eq}}\},\{\mathbf{r}\})$. An analogy may be drawn between the ground state Born–Oppenheimer potential energy and $U_g(\{\mathbf{r}\})$. The charges $\{q\}$ in this case play a role similar to the electronic wave functions, and they represent the electronic degrees of freedom. In fact, Eq. (6) is the principle of chemical potential equalization[18] in the density functional electronic structure theory[19] recast in the point-charge representation. In the adiabatic limit, the nuclei evolve dynamically on the Born–Oppenheimer potential energy surface. In the same fashion, the principle of electronegativity equalization requires the charge density to fluctuate adiabatically accordingly as the nuclear coordinates evolve in time. One can obtain the same

set of coupled linear equations by variationally minimizing the total energy $U(\{q\},\{\mathbf{r}\})$ with respect to the charge distribution, subject to the constraints of molecular charge conservation,

$$\sum_\alpha q_{i\alpha} = 0, \tag{7}$$

or the constraint of overall charge conservation,

$$\sum_i \sum_\alpha q_{i\alpha} = 0. \tag{8}$$

(We have formulated the Lagrangian equations for neutral molecules, however, extension to molecular ions is trivial.) Note that Eqs. (7) and (8) are holonomic constraints, and the redundant variables in $\{q\}$ may be eliminated by a transformation to a set of ''generalized'' variables, denoted by $\{Q\}$ [$N_{\text{mole}}N_{\text{site}} - N_{\text{mole}}$ for the constraint in Eq. (7)]:

$$\mathbf{q} = \mathbf{C}^\dagger \mathbf{Q}, \tag{9}$$

where $\{Q\}$ and $\{q\}$ are represented collectively by column vectors $\mathbf{Q}$ and $\mathbf{q}$; $\mathbf{C}$ is an $(N_{\text{mole}}N_{\text{site}})$ by $(N_{\text{mole}}N_{\text{site}} - N_{\text{mole}})$ matrix, and $\mathbf{C}^\dagger$ denotes the transpose matrix of $\mathbf{C}$. Thus, Eq. (4) becomes

$$U_{ee}(\{\mathbf{r}\},\{q\}) = \mathbf{Q}^\dagger \mathbf{C}\mathbf{X} + \frac{1}{2}\mathbf{Q}^\dagger \mathbf{C}\mathbf{J}\mathbf{C}^\dagger \mathbf{Q}, \tag{10}$$

where $\{\chi\}$ and $\{J\}$ are represented collectively by an $N_{\text{mole}}N_{\text{site}}$-dimension column vector, $\mathbf{X}$, and an $N_{\text{mole}}N_{\text{site}}$ by $N_{\text{mole}}N_{\text{site}}$ matrix, $\mathbf{J}$, respectively. The requirement of electronegativity equalization [Eq. (6)] thus leads to

$$\mathbf{C}\mathbf{J}\mathbf{C}^\dagger \mathbf{Q} = -\mathbf{C}\mathbf{X}. \tag{11}$$

It can be easily shown that

$$U_{ee}^{\text{eq}} = U_{ee}(\{q^{\text{eq}}\},\{r\}) = \frac{1}{2}\sum_i \sum_\alpha \chi_{i\alpha}^0 q_{i\alpha}^{\text{eq}} = \frac{1}{2}\mathbf{q}^{\text{eq}\dagger}\mathbf{X}, \tag{12}$$

$$\mathbf{q}^{\text{eq}} = -\mathbf{C}^\dagger(\mathbf{C}\mathbf{J}\mathbf{C}^\dagger)^{-1}\mathbf{C}\mathbf{X} = -\mathscr{T}\mathbf{X}, \tag{13}$$

where $\mathscr{T}$ denotes $\mathbf{C}^\dagger(\mathbf{C}\mathbf{J}\mathbf{C}^\dagger)^{-1}\mathbf{C}$. When an *external* potential, $v(\mathbf{r})$, is applied to the system, the charges redistribute so as to equalize the electronegativity at each charge sites. Since the external potential couples to the system *linearly* [i.e., the total energy defined in Eq. (3) includes a term, $\Sigma q_{i\alpha}v(\mathbf{r}_{i\alpha})$], Eq. (11) becomes

$$\mathbf{C}\mathbf{J}\mathbf{C}^\dagger \mathbf{Q}^F = -C(\mathbf{X}+\mathbf{v}), \tag{14}$$

where the column vector $\mathbf{v}$ represents collectively the external potential applied to each charge sites, $\{v(\mathbf{r})\}$. The above equation suggests that the polarization response, $\Delta\mathbf{Q}$, defined here as the change in charge distribution upon application of an external electric field, satisfies

$$\mathbf{C}\mathbf{J}\mathbf{C}^\dagger \Delta\mathbf{Q} = -\mathbf{C}\mathbf{v}, \tag{15}$$

when the principle of electronegativity equalization is applied in the presence of an external electric field. The ground-state charges and the total energy in this case become

$$U_{ee}^F = U_{ee}(\{q^F\},\{\mathbf{r}\}) = \frac{1}{2}\sum_i \sum_\alpha [\chi_{i\alpha}^0 + v(\mathbf{r}_{i\alpha})]q_{i\alpha}^F, \tag{16}$$

$$\mathbf{q}^F = -\mathscr{T}(\mathbf{X}+\mathbf{v}) = \mathbf{q}^{\text{eq}} - \mathscr{T}\mathbf{v}, \tag{17}$$

$$\Delta\mathbf{q} = -\mathscr{T}\mathbf{v}. \tag{18}$$

From Eqs. (12)–(18), Eq. (16) can be rewritten as

$$U_{ee}^F = U_{ee}^{\text{eq}} + \frac{1}{2}\mathbf{v}^\dagger \mathbf{q}^{\text{eq}} - \frac{1}{2}\mathbf{v}^\dagger\mathscr{T}\mathbf{X} - \frac{1}{2}\mathbf{v}^\dagger\mathscr{T}\mathbf{v}$$

$$= U_{ee}^{\text{eq}} + \mathbf{v}^\dagger \mathbf{q}^{\text{eq}} - \frac{1}{2}\mathbf{v}^\dagger\mathscr{T}\mathbf{v}. \tag{19}$$

It is apparent from the above equation that the total energy of the system is the sum of the total energy free of an external field, the interaction energy between the ''free'' system and the external field, and the polarization energy. If the applied external potential corresponds to a spatially uniform electric field, i.e.,

$$v(\mathbf{r}) = -\mathbf{r}\cdot\mathbf{E}, \tag{20}$$

the total energy of the system is simply [see also, for example, Eq. (4.8) of Ref. 11]

$$U_{ee}^F = U_{ee}^{\text{eq}} - \sum_i^{N_{\text{mole}}} \sum_\alpha^{N_{\text{site}}} q_{i\alpha}^{\text{eq}}\mathbf{r}_{i\alpha}\cdot\mathbf{E}$$

$$- \frac{1}{2}\sum_{i,j}^{N_{\text{mole}}} \sum_{\alpha,\beta}^{N_{\text{site}}} \mathbf{E}\cdot\mathbf{r}_{i\alpha}\mathscr{T}_{i\alpha j\beta}\mathbf{r}_{j\beta}\cdot\mathbf{E}. \tag{21}$$

Thus, we find that the quadratic FQ potential function gives rise to a linearly polarizable system, with the polarizability given by

$$\mathbf{A} = \sum_{i,j} \sum_{\alpha,\beta} \mathscr{T}_{i\alpha j\beta}\mathbf{r}_{i\alpha}\otimes\mathbf{r}_{j\beta}, \tag{22}$$

where $\otimes$ denotes a direct product. Equation (22) appears to suggest that the polarizability would depend on the origin of the coordinate system. Such coordinate dependency is removed by the structure of the $\mathscr{T}$ matrix. In addition, when all the charge sites are in the same plane, the out-of-plane polarizability vanishes. This result may be obtained by a coordinate transformation to place the system in, for example, the x-y plane (i.e., $z=0$). The zz component of the polarizability tensor thus becomes

$$\mathbf{A}_{zz} = \text{const}\sum_{i,j}^{N_{\text{mole}}} \sum_{\alpha,\beta}^{N_{\text{site}}} \mathscr{T}_{i\alpha j\beta}, \tag{23}$$

and it vanishes because

$$\sum_{i,j}^{N_{\text{mole}}} \sum_{\alpha,\beta}^{N_{\text{site}}} \mathscr{T}_{i\alpha j\beta} = 0, \tag{24}$$

which is apparent by noting that there would be no charge transfer; that is, all charges are equal to zero if $\chi_{i\alpha}^0$ are identical in Eq. (13).

As discussed above, the point charges $\{q\}$ represent the electronic degrees of freedom, just as wave functions do in molecular orbital theory. However, while both the Hartree–

Fock equation in *ab initio* molecular orbital theory and the Kohn–Sham equation in the density functional theory are nonlinear, and must be solved by iterative self-consistent procedures, the corresponding counterpart in the FQ model, Eq. (6), comprises a set of coupled *linear* equations. This is because the total energy functional defined in Eq. (3) is quadratic in $\{q\}$.

In the empirical FQ potential of Rick, Stuart, and Berne,[11] $\chi_{i\alpha}^0$ in Eq. (3) is set equal to $\tilde{\chi}_\alpha^0$ and is independent of $\{\mathbf{r}\}$; the absolute hardness $J_{\alpha\alpha}^0$ and the intramolecular screened Coulomb interaction, $J_{i\alpha i\beta}$, are evaluated analytically as the two-electron Coulomb integral between two s-type Slater orbitals,[20]

$$J_{i\alpha i\beta}(\mathbf{r}_{i\alpha i\beta}) = \int d\mathbf{r}_a d\mathbf{r}_b \phi_{i\alpha}^2(\mathbf{r}_a) \frac{1}{|\mathbf{r}_a - \mathbf{r}_b|} \phi_{i\beta}^2(\mathbf{r}_b), \quad (25)$$

$$\phi_{i\alpha}(\mathbf{r}_a) = N_\alpha |\mathbf{r}_a - \mathbf{r}_{i\alpha}|^{n_\alpha - 1} e^{-\zeta_\alpha(\mathbf{r}_a - \mathbf{r}_{i\alpha})}, \quad (26)$$

where $\mathbf{r}_a$ and $\mathbf{r}_b$ are the coordinates of the electrons; the intermolecular Coulomb interaction $J_{i\alpha j\beta}$ is approximated by $1/|\mathbf{r}_{i\alpha} - \mathbf{r}_{j\beta}|$. The term $V(r_{i\mu j\nu})$ in Eq. (3) represents the pair interaction, which was taken to be a Lennard-Jones interaction between oxygens in the empirical FQ potential. In the empirical FQ potential, the geometry of the charge sites is taken to be that of the TIP4P model; the Slater exponents, the values for $\tilde{\chi}_\alpha^0$, and the Lennard-Jones parameters are adjusted to reproduce the gas-phase monomer dipole, and to achieve optimal agreement with experiment regarding properties such as the energy, pressure, and pair-correlation functions of the liquid phase.

III. NUMERICAL METHODS

A. *Ab initio* calculations

Ab initio calculations are performed for 57 water trimers and 94 dimers using the PSGVB program.[21] The total interaction energy for an N_{mole}-molecule system is, in general, decomposed into components due to one body, two bodies, three bodies,...,etc.:

$$U^{(1)} = \sum_{i=1}^{N_{\text{mole}}} U(w_i), \quad (27)$$

$$U^{(2)} = \sum_{i=1}^{N_{\text{mole}}} \sum_{j>i}^{N_{\text{mole}}} U(w_i w_j) - 2 U^{(1)}, \quad (28)$$

$$U^{(3)} = \sum_{i=1} \sum_{j>i} \sum_{k>j>i} U(w_i w_j w_k) - U^{(2)} - U^{(1)}, \quad (29)$$

$$\dots,\text{etc.},$$

where w_i, $w_i w_j$, $w_i w_j w_k$, etc., denote water monomers, dimers, trimers, etc., respectively; subscripts i, j, and k label the individual water molecules that make up a cluster; the functional form of U in Eqs. (27)–(29) depends on the model [i.e., FQ, as defined in Eq. (3), or *ab initio* energetics], and the argument of U in Eqs. (27)–(29) implies that it is evaluated for one, two, and three water molecules,..., etc., respectively. The two-body interaction energy is evaluated for 94 pairs of water molecules ($N_{\text{mole}} = 2$) using the local

Møller–Plesset second-order perturbation (LMP2) theory[22–24] and Dunning's correlation-consistent triple-zeta basis function, cc-pVTZ(-f) basis set[25–28] ($3s2p1d$ for hydrogen and $4s3p2d$ for oxygen):

$$U_{\text{HF/LMP2}}^{(2)} = U_{\text{HF/LMP2}}(w_1 w_2) - U_{\text{HF/LMP2}}(w_1)$$
$$- U_{\text{HF/LMP2}}(w_2) + U_{\text{CP}}^{(2)}, \quad (30)$$

where $U_{\text{CP}}^{(2)}$ denotes counterpoise corrections,[29]

$$U_{\text{CP}}^{(2)} = H_{\text{HF}}(w_1) + U_{\text{HF}}(w_2) - U_{\text{HF}}(w_1\{w_2\})$$
$$- U_{\text{HF}}(w_2\{w_1\}), \quad (31)$$

in which $U_{\text{HF}}(w_i)$ is the Hartree–Fock energy of monomer w_i calculated with monomer w_i's basis functions only, while $U_{\text{HF}}(w_i\{w_j\})$ is the Hartree–Fock energy of monomer w_i calculated using the basis functions for the complex $w_i w_j$.

The three-body interaction energy is calculated at the Hartree Fock level using the 6-31G** basis set,

$$U_{\text{HF}}^{(3)} = U(w_1 w_2 w_3) - U_{\text{HF}}^{(2)} - U_{\text{HF}}^{(1)}. \quad (32)$$

$U_{\text{HF}}^{(2)}$ in Eq. (32) is calculated in the same way as $U_{\text{HF/LMP2}}^{(2)}$ in Eq. (31), except that $U_{\text{HF}}^{(2)}$ in Eq. (32) is calculated at the Hartree Fock level and that the basis functions of the trimers are used for the counterpoise correction. Likewise, $U_{\text{HF}}^{(1)}$ in Eq. (32) is calculated using the trimer basis functions. We also calculated the three-body energies at the LMP2 level using the cc-pVTZ(-f) basis set for a selection of 23 out of the 57 trimers, and the results of 6-31G** basis set at the Hartree Fock level are systematically less attractive (i.e., less negative) than the results at the LMP2 level using the cc-pVTZ(-f) basis set, but they are in agreement to within 0.3 kcal/mol. The higher-order many-body terms (i.e., $U^{(n)}$ with $n \geq 4$) are found to be negligible,[30] and are therefore not considered in the present modeling of force fields.

B. Extended Lagrangian method

In the preceding section, we have formulated the FQ model in a compact matrix representation, and a matrix inversion is required to solve for the partial charge carried by each atom in the system. Matrix inversion, however, is a difficult proposition, if not impossible, for large-scale molecular dynamics simulations. A more efficient strategy is to treat the electronic degrees of freedom along with the nuclear degrees of freedom as dynamical variables, and to use the so-called "extended Lagrangian method."[31–34] This method relies on the presumption that the electronic degrees of freedom evolve adiabatically with respect to the change in the nuclear degrees of freedom. The extended Lagrangian corresponding to the energy defined in Eq. (3) is

$$\mathscr{L} = \frac{1}{2} \sum_{i=1}^{N_{\text{mole}}} \left(\sum_{\mu=1}^{N_{\text{atom}}} M_\mu \dot{r}_{i\mu}^2 + \sum_{\alpha=1}^{N_{\text{site}}} m_q \dot{q}_{i\alpha}^2 \right) - U(\{\mathbf{r}\}, \{q\})$$
$$- \sum_{i=1}^{N_{\text{mole}}} \lambda_i \sum_{\alpha=1}^{N_{\text{site}}} q_{i\alpha} - \sum_{\gamma=1}^{N_{\text{bond}}} \Gamma_\gamma G_\gamma(\{\mathbf{r}\}), \quad (33)$$

where M_μ is the mass of atom type μ, and m_q is the fictitious mass associated with the charges $\{q\}$. It does not correspond to any physical mass and is simply set to a value small

enough such that the charges follow the atomic coordinates adiabatically. The Lagrangian defined in Eq. (33) also includes an N_{bond} number of rigid bonding constraints, $G_\gamma(\{\mathbf{r}\})$, and an N_{mole} number of constraints to ensure that each molecule remains electrostatically neutral. Readers should refer to Ref. 11 for a Lagrangian that allows intermolecular charge transfer, and to Ref. 35 for a discussion on the problematic and unphysical charge transfer between atoms that are far apart.[18]

Based on the Lagrangian defined in Eq. (33), the equations of motion for the positions and charges are

$$M_{i\mu}\ddot{r}_{i\mu} = -\frac{\partial}{\partial r_{i\mu}}\left[U(\{\mathbf{r}\},\{q\}) + \sum_{\gamma=1}^{N_{\text{bond}}}\Gamma_\gamma G_\gamma(\{\mathbf{r}\})\right], \quad (34)$$

and

$$m_q\ddot{q}_{i\alpha} = -\frac{\partial U(\{\mathbf{r}\},\{q\})}{\partial q_{i\alpha}} - \lambda_i. \quad (35)$$

These equations of motion can be integrated by many standard schemes. The formulation here is for a constant energy, constant volume ensemble; other extended Lagrangians can be formulated for simulations at constant temperature and constant pressure.

An alternative to the extended Lagrangian method for solving Eq. (6) is the iterative procedure.[36,37] Various attempts have been made to improve the accuracy, stability, as well as efficiency of this approach.[38–40] Although a direct comparison between the iterative procedure and the extended Lagrangian method is lacking, we found the two methods to be comparable in terms of computational speed based on indirect evidence from the literature. We note that on a massively parallel computer such as a CM5, the extended Lagrangian method is roughly 20 times faster than the standard matrix inversion procedure. According to the benchmark calculation of Bernardo *et al.*[41] on CM5 using a different polarizable water potential, the iterative method is roughly 17 times faster than direct matrix inversion to achieve a convergence of 1.0×10^{-8} D in the induced dipole.

C. Parametrization and simulations

A nonlinear Newton–Raphson optimization scheme is used to search for a set of potential function parameters that give the optimal fit to the *ab initio* data. This method is also used to obtain the dimer and trimer minimum energy configurations for the empirical and the new *ab initio* FQ potentials.

Unless otherwise noted, all the intramolecular $J_{i\alpha i\beta}$ are calculated using Eq. (25), and all the intermolecular $J_{i\alpha j\beta}$ are calculated as $1/r_{i\alpha j\beta}$. All the molecular dynamics simulations are performed on CM5 for the NVE ensembles after equilibrating their temperature to 298 K. The parameters (e.g., the number of terms included in the Ewald sum, etc.) and methods for molecular dynamics simulations are the same as in Ref. 11, unless otherwise noted. To calculate the frequency-dependent dielectric constants, we use Filon's method for Fourier–Laplace transformation.[42,43]

IV. MODELING MANY-BODY INTERACTIONS

In this section, we examine the many-body interactions exhibited in water trimers, and the polarization energy of a monomer in the presence of spatially uniform and nonuniform external fields, respectively. In particular, we look for their similarity and differences within the framework of the Fluctuating Charge Model [Eq. (4)].

Following Eq. (19), the polarization energy of any system with respect to an external electric potential in the FQ model is

$$E_{\text{pol}}^{\text{FQ}} = -\frac{1}{2}\,\mathbf{v}^\dagger\mathscr{T}\mathbf{v}. \quad (36)$$

For a water molecule, it can be explicitly written as

$$E_{\text{pol}}^{\text{FQ}} = -\frac{C_{zz}}{4}(v_{\text{H}a} + v_{\text{H}b} - 2v_{\text{O}})^2 - \frac{C_{yy}}{4}(v_{\text{H}a} - v_{\text{H}b})^2, \quad (37)$$

where

$$C_{zz} = [2J_{\text{OO}}^0 + J_{\text{HH}}^0 - 4J_{\text{OH}}(r_{M\text{H}}) + J_{\text{HH}}(r_{\text{HH}})]^{-1}, \quad (38)$$

$$C_{yy} = [J_{\text{HH}}^0 - J_{\text{HH}}(r_{\text{HH}})]^{-1}, \quad (39)$$

and v_α is the external electric potential $v(\mathbf{r})$ applied at the αth charge site, $\mathbf{r} = \mathbf{r}_\alpha$. If the external field is uniform, i.e.,

$$v(\mathbf{r}) = -\mathbf{E}\cdot\mathbf{r}, \quad (40)$$

the FQ polarization energy in Eq. (37) becomes

$$E_{\text{pol}}^{\text{FQ}} = -\frac{1}{2}\,\alpha_{zz}|\mathbf{E}\cdot\mathbf{u}_z|^2 - \frac{1}{2}\,\alpha_{yy}|\mathbf{E}\cdot\mathbf{u}_y|^2, \quad (41)$$

where $\mathbf{u}_z$ and $\mathbf{u}_y$ are the unit vectors along the C_2 axis and along the direction perpendicular to the C_2 axis in the molecular plane; and α_{zz} and α_{yy} are

$$\alpha_{zz} = 2z_{M\text{H}}^2 C_{zz}, \quad (42)$$

$$\alpha_{yy} = \frac{1}{2}\,r_{\text{HH}}^2 C_{yy}, \quad (43)$$

with the $\mathbf{u}_z$ component of the distance from the M site to the hydrogen denoted by $z_{M\text{H}}$ and the distance between the two hydrogens denoted by r_{HH}. It is apparent from Eq. (41) that α_{zz} and α_{yy} correspond to the polarizabilities in the molecular plane in the presence of a uniform external field.

The number of independent variables N_v in the fitting of polarization energy is, in general,

$$N_v = \frac{1}{2}\left\{N_{\text{site}}^2 - (2N_{\text{site}} - 1)\sum_p^{N_p} N_{ip}\right.$$
$$\left. + \sum_p^{N_p} N_{ip}^2 + 2N_p N_{\text{site}} - N_{\text{site}} - 2N_p\right\}, \quad (44)$$

where N_{ip} denotes the number of charge sites that are equivalent under symmetry operation and N_p denotes the total number of such sets. For example, the two charge sites on the hydrogen atoms in the water molecule are identical under symmetry operation. Thus N_{i1} and N_p equal 2 and 1, respectively, and there are two independent variables avail-

able for fitting the polarization energy of a water molecule. (Note that in the empirical FQ potential of Rick *et al.*, the Slater orbital exponents of oxygen and hydrogen are varied to fit the empirical data. Since both the J_{OO}^0 and J_{OH} (r_{MH}) are calculated using the two-electron Coulomb integral in the empirical FQ potential of Rick *et al.* [see Eq. (25)], this additional constraint may make it impossible to obtain the optimal α_{zz} with any value of ζ_O. Thus, a more flexible parameter search is to vary, for example, the value of ζ_O for the one-center two-electron integral while keeping the value of ζ_O for the two-center two-electron integral fixed.)

In the following, we calculate the polarization energy of the FQ model [Eq. (36)] in the presence of a nonuniform external field, and the three-body energy in water trimers [Eq. (29)] using various sets of values of the Slater exponents ζ_O and ζ_H, which, in turn, specify the values of α_{yy} and α_{zz}, the polarizabilities along the HH axis and the C_2 axis.

A. Polarization due to nonuniform electric field

If many-body interactions could be approached by monomer properties, the simplest model would be a potential function that accounts for the dipole polarization. In the case of the fluctuation charge model, one would then derive the potential function parameters by fitting the molecular polarizabilities to the *ab initio* values. This is equivalent to fitting the polarization energy of a fluctuating charge model to the corresponding *ab initio* polarization energy due to a uniform external field. The instantaneous local electric field experienced by a molecule in the condensed phases is, however, not uniform. And it is not unlikely that distinct sets of potential function parameters would be obtained when the *ab initio* polarization energy due to external uniform and nonuniform electric fields are used in the parametrization. This question is addressed below.

The molecular polarizability of the water monomer, α_{xx}^{QM}, α_{yy}^{QM}, and α_{zz}^{QM}, are estimated to be 0.7464, 1.1833, and 0.9915 Å^3, respectively, using the cc-pVTZ(-f) basis at the Hartree–Fock level. The *ab initio* polarization energy of a water molecule in a nonuniform field due to a probe charge is calculated as follows. When a probe charge is placed around a water molecule, it gives rise to a perturbation, V^{ext}, and the *ab initio* polarization energy E_{pol}^{QM} is calculated as the difference between $\langle \psi^F | V^{ext} | \psi^F \rangle$ and $\langle \psi^{vac} | V^{ext} | \psi^{vac} \rangle$, which are the expectation values of V^{ext} evaluated in terms of the wave functions in equilibrium with the external charge (ψ^F) and in vacuum (ψ^{vac}), respectively. By varying the location of the probe charge relative to the water molecule, and keeping the probe charge always at least 2 and 2.5 Å away from the hydrogen and oxygen, respectively, we calculate E_{pol}^{QM} for a set of 1034 positions of the probe charge relative to the water molecule. The probe charge is either $-0.2e$ or $+0.2e$.

Figure 1 shows the correlation between the polarizabilities and δ_{pol}, which measures the agreement between E_{pol}^{FQ} and E_{pol}^{QM}, and is defined as the total absolute difference between E_{pol}^{FQ} and E_{pol}^{QM} for the 1034 geometries (in kcal/mol). This quantity is plotted against the polarizability, α_{yy} and α_{zz} [panels (a) and (b), respectively]. The variation in the polarizability along one direction is made while keeping the
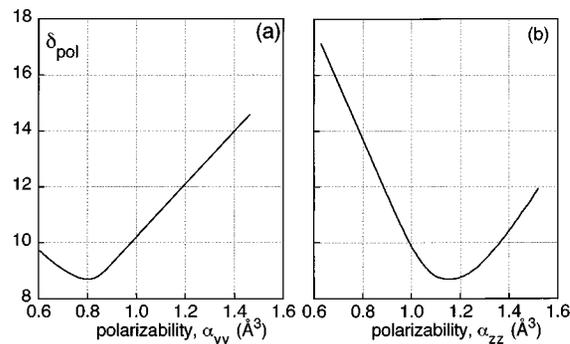

FIG. 1. The correlation between δ_{pol} (in kcal/mol) and the molecular polarizabilities, α_{yy} and α_{zz} (Å^3) [panels (a) and (b), respectively]. The quantity δ_{pol} is the total absolute difference between *ab initio* polarization energy (E_{pol}^{QM}) and that of the FQ models (E_{pol}^{FQ}) for a set of 1034 configurations of the external probe charge of ($\pm 0.2e$). This quantity is a measure of the agreement between E_{pol}^{QM} and E_{pol}^{FQ}. In panel (a), α_{yy} is varied while keeping α_{zz} fixed at a value of 1.1587 Å^3. In panel (b) α_{zz} is varied and α_{yy} is fixed at a value of 0.8067 Å^3.

other constant. Figure 1 shows that the best agreement between E_{pol}^{FQ} and E_{pol}^{QM} is obtained when α_{yy}^{FQ} and α_{zz}^{FQ} are 0.8067 and 1.1587 Å^3, respectively. Compared to the *ab initio* results, of α_{yy}^{FQ} is reduced by nearly 32% and α_{zz}^{FQ} increased by 17%. Since the out-of-plane polarization is not present in the point-charge FQ model, it is tempting to attribute the large difference between α_{yy}^{FQ} and α_{yy}^{QM} to the inclusion of E_{pol}^{QM} due to out-of-plane probe charges in the fitting. We found that this difference is reduced by only about 11% when the out-of-plane configurations are excluded from the fitting, and thus it cannot be sufficiently accounted for by this line of reasoning. It may also be suggested that the electric fields created by the probe charges are beyond the range for which the linear response is valid. In Fig. 2 we plot the actual *ab initio* polarization energy of water with respect to the field strengths (data points) and the expected linear polarization response (lines) using the *ab initio* molecular polarizabilities listed in the preceding section. The field strength due to a probe charge of $\pm 0.2e$ at a distance of 2.5 Å away is marked by a thin solid vertical line. It is clear from
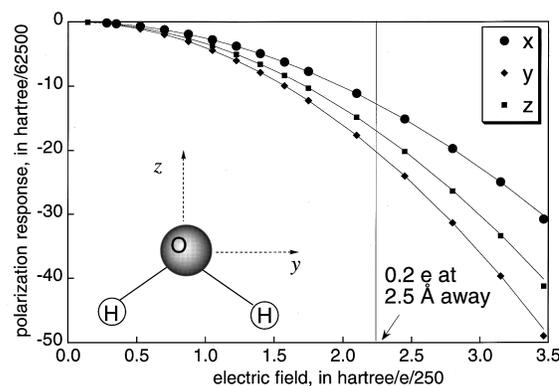

FIG. 2. The *ab initio* polarization energy (E^{pol}) of a water monomer responding to homogeneous external electric fields applied along the x, y, and z directions are represented by circles, diamonds, and squares, respectively. The dashed line marks the magnitude of the electric field created by a point charge of $0.2e$ at 2.5 Å away.

Fig. 2 that the polarization response is nearly linear (with a very slight deviation) in the presence of uniform fields with strengths up to 0.014 Hartree/e. Our results presented above thus show that two distinct sets of α's (and therefore two distinct sets of potential parameters) are obtained, depending on whether the polarization energy due to uniform or the nonuniform external fields is used for parametrization.

B. Three-body energies of water trimers

In this section, we explore another approach for modeling many-body interactions. In this approach, many-body interactions are assumed to form a converging sequence of energy terms involving three bodies, four bodies,..., etc. If this sequence converges within a few terms involving a small number of molecules, it is then possible to extract the essence of many-body interactions from studies on molecular clusters. Using *ab initio* quantum chemical methods, the many-body interactions of water clusters are first partitioned into components corresponding to the two-body, three-body,..., etc., and then the FQ potential parameters are adjusted to reproduce the *ab initio* many-body energies. Prior *ab initio* calculations have shown that the many-body energies involving more than three water molecules are negligible, and therefore *ab initio* quantum chemical calculations on water trimers will suffice for analyses and parametrization. (Note that this is not to say that the ''binding energy'' of water clusters larger than trimers is negligible, but that the contributions to the total binding energy from many-body energy terms [see Eq. (29)] involving more than three water molecules are insignificant.) We analyze the interactions of water trimers in various relative orientations, and give a qualitative description of the origin of the three-body energies of water trimers. Finally, we seek a correlation between the three-body energies and the molecular polarizabilities.

Three-body interaction energies calculated for 57 trimers using Eq. (32), i.e., at the Hartree–Fock (HF) level are summarized below. The geometry of the 57 trimers are randomly selected from liquid simulations and in part selected to include geometries with bifurcated hydrogen bonds. The monomers are constrained to be nearly the equilibrium water monomer geometry, with the largest deviation of 0.015 Å in the OH bond length and 3.15° in the HOH bond angles. The three-body energies vary from very negative (attractive, $\sim$ −1.18 kcal/mol) to positive (repulsive, $\sim$0.83 kcal/mol). The magnitude of the three-body interaction can be as much as 10% of the overall binding energy of a trimer. Further analysis shows that the trimers with a repulsive three-body energy greater than 0.2 kcal/mol involve bifurcated hydrogen bonds, with one of the water molecules either accepting a hydrogen from each of the remaining two water molecules or donating its two hydrogens to both of the other two water molecules at the same time. An example of the bifurcated hydrogen-bonded configuration is given in Fig. 3. The repulsive three-body interaction exhibited in these bifurcated hydrogen-bonded trimers may at first seem unexpected based on the Mulliken analysis as well as electrostatic potential mapping of the charge distribution of a hydrogen-bonded water dimer. It is found in a water dimer that the oxygen becomes more negatively charged, or equivalently, more
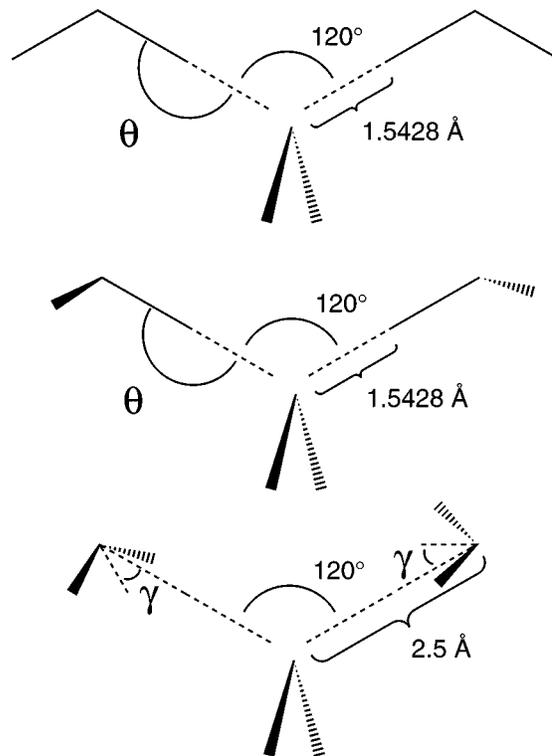

FIG. 3. A representative structure for bifurcated hydrogen-bonded trimers.

''basic'' upon accepting a hydrogen bond from the other water molecule.[44] One therefore expects that it is energetically favorable to accept a second hydrogen bond from the third water molecule, and this would lead to an *attractive* three-body energy, contradicting to the results of *ab initio* quantum chemical calculations. This paradox may be resolved by considering the quantum-mechanic character of hydrogen bonding. The repulsive three-body interaction reflects the unfavorable molecular orbital interaction for the oxygen to form a second hydrogen bond. In contrast to electrostatic forces, this kind of molecular orbital interaction is expected to be short ranged. And, indeed, prior *ab initio* calculations show a rapid decay of such repulsive three-body interaction as the distance between the monomers becomes larger.[8,45] In the present study, we also examine the orientational dependence of the repulsive three-body energies, and the results are summarized in Table I.

To test the convergence of the *ab initio* three-body energies with respect to the level of theory, we also applied the local Møller–Plesset second-order perturbation (LMP2) method to a subset of 23 trimers (also including counterpoise corrections), and the results are compared to the HF results

TABLE I. Orientational dependence of the *ab initio* three-body energy. The geometrical parameters θ and γ are defined in Fig. 3.

θ	Structure I	Structure II	γ	Structure III
180	1.99	2.02	0	0.58
150	2.00	1.45	52.3	2.02
120	2.46	1.58	82.3	0.78
			104.5	0.33

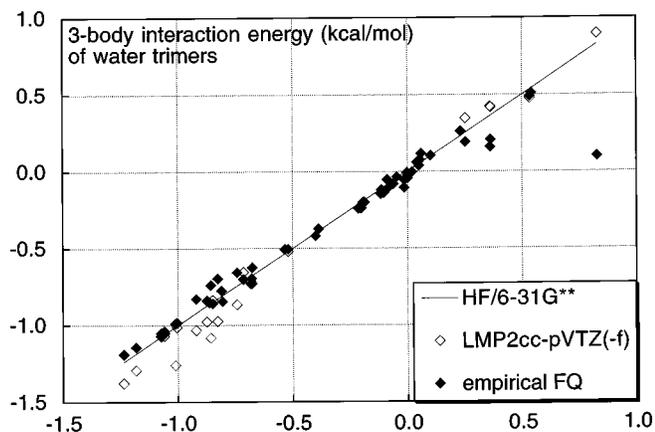J. Chem. Phys., Vol. 108, No. 12, 22 March 1998

Liu *et al.*   4747

FIG. 4. Three-body interaction energies (kcal/mol) of 57 water trimers calculated by the empirical FQ potential of Rick *et al.* (solid diamonds) and by *ab initio* quantum chemistry at the local MP2 level [cc-pVTZ(-f) basis set] (open diamonds) are plotted against the the *ab initio* results at the Hartree–Fock level (6-31G ** basis set). The solid line representing perfect agreement with the HF results is also plotted.

FIG. 5. The correlation between δ_{3b} (in kcal/mol) and the molecular polarizabilities, α_{yy} and α_{zz} (in Å^3) [panels (a) and (b), respectively]. The quantity δ_{3b} is the total absolute difference between *ab initio* three-body energy (E_{3b}^{QM}) and that of the FQ models (E_{3b}^{FQ}) for a subset of 43 water trimers, all with negative (i.e., attractive) three-body energy. This quantity, δ_{3b}, is a measure of the agreement between E_{3b}^{QM} and E_{3b}^{FQ}. The solid lines are calculated with α_{zz} [panel (a)] and α_{yy} [panel (b)], fixed at the same value as in panels (a) and (b) of Fig. 1, respectively. The dashed line in panels (a) and (b) is calculated with α_{zz} fixed at 0.9714 and α_{yy} at 0.9062 Å^3, respectively.

in Fig. 4. The solid line represents perfect agreement with the HF calculations. The LMP2 results are represented by open diamonds. As seen in Fig. 4, the HF and LMP2 results are nearly indistinguishable, suggesting that there are negligible three-body effects in the correlation energy and therefore accurate three-body energy can be obtained at the HF level.

Also presented in Fig. 4 are the three-body interaction energies for the same 57 trimer configurations calculated using the empirical FQ potential [Eq. (29)] (solid diamonds). Except for a number of trimers, the FQ model gives a remarkably good prediction of three-body energies, and the agreement with the *ab initio* results is within 0.3 kcal/mol (i.e., within the uncertainty introduced by the type of basis sets and the level of *ab initio* methods). The exceptions are those that involve bifurcated hydrogen bonds shown in Fig. 3. As discussed in the beginning of this section, repulsive three-body energy observed in this type of bifurcated hydrogen-bonded trimers contradicts the usual electrostatic model of hydrogen bonding. The result in Fig. 4 indicates that the functional form of the empirical FQ potential is adequate for depicting many-body interactions when electrostatic polarization is the predominant source of many-body interactions, but not adequate for the quantum-mechanical many-body effects, as exhibited in the bifurcated hydrogen-bonded water trimers.

It is also worth noting that the intermolecular Coulombic interaction, $J_{i\alpha j\beta}$, may be modeled by the two-electron Coulomb integral rather than by $1/r_{i\alpha j\beta}$. We found that in this case, the three-body energies are systematically less attractive than the *ab initio* results, but the agreement is still within 0.3 kcal/mol, except for the bifurcated hydrogen-bonded trimers. When intermolecular charge transfer is allowed, a significant deviation from the *ab initio* results is observed. This is due to the unphysical charge transfer between water molecules that are far apart, and we found that this problem cannot be remedied by fitting the potential function parameters to the *ab initio* three-body energies.
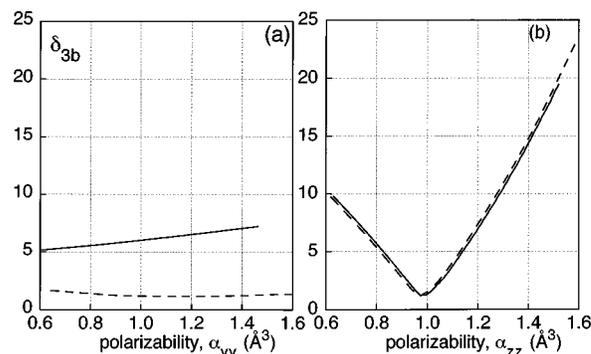
Finally, to further explore the possibility of approaching many-body interactions by monomer properties, we seek a correlation between the molecular polarizabilities and the three-body energy of water trimers. Specifically, for those water trimers exhibiting attractive three-body interactions, the ability of a set of FQ parameters to reproduce faithfully their *ab initio* three-body energy should depend on how well they represent the molecular polarization response of a water molecule to a nonuniform external field. We define the agreement between E_{3b}^{FQ} and E_{3b}^{QM} as the total absolute difference between E_{3b}^{FQ} and E_{3b}^{QM} for the 43 water trimers, all with negative (i.e., attractive) three-body energies, and denote it by δ_{3b} (in kcal/mol). This quantity is plotted against the polarizabilities, α_{yy} and α_{zz} [panels (a) and (b), respectively], in Fig. 5. The variation in the polarizability along one direction is made while keeping the other constant. One finds that δ_{3b} correlates only with α_{zz}, the polarizability along the C_2 axis direction. This is in sharp contrast with the correlation between δ_{pol} and the molecular polarizabilities (cf. Fig. 1). Figure 5 suggests that the three-body energies in our parametrization database are largely determined by α_{zz}, but not dependent upon the polarizability along the HH axis direction. This result shows that one would obtain distinct sets of molecular polarizabilities (and therefore distinct sets of FQ potential parameters), depending on whether the *ab initio* polarization energy due to nonuniform fields or the *ab initio* three-body energies of water trimers are used for parametrization.

V. *AB INITIO* DYNAMICALLY FLUCTUATING CHARGE MODEL

In the preceding section we analyzed the two different approaches for modeling many-body interactions and we found interesting differences between them. These differences merit further studies. To explore other important issues relevant to constructing *ab initio* force fields (as mentioned in the Introduction), we choose to parametrize the many-body interactions by fitting to the three-body energies of wa-

ter trimers. We do so because our own analyses and also prior *ab initio* quantum chemical calculations show that the three-body energies calculated at the Hartree–Fock level with the cc-pVTZ(-f) basis set are sufficiently converged. The question of convergence of the polarization response to an electric field is far more complicated, and has thus far been examined primarily for the case of a uniform external electric field. In order to obtain quantitatively accurate molecular polarizabilities (i.e., within a few percent of the experimental results), our results[46] and those of others[47] indicate that it is necessary to utilize diffuse functions and at least the MP2 level of electron correlation. This is a significantly higher level of theory than is required, at least in the case of water trimers, to converge the three-body energy. The disparity in the convergence behavior of molecular polarizabilities and many-body energy suggests that there are components of the quantum mechanical polarization response that have little impact upon many-body energies. The investigation of this issue, the convergence of the polarization response in nonuniform external fields with respect to level of quantum chemical calculations, and the implications for developing methods for fitting polarizable models such as the FQ model to the *ab initio* response data (as opposed to three-body energies, which are significantly more cumbersome to calculate), will be pursued further in another paper.

The specific procedures for deriving the *ab initio* many-body water potential is as follows. The potential parameters pertinent to many-body interactions are adjusted to reproduce *ab initio* three-body energy of water trimers. In the present case, these parameters are the potential parameters of the FQ model, the χ's and the ζ's. The values of these parameters are then held fixed, and the potential parameters corresponding for interactions between pairs of molecules are then adjusted to give the optimal agreement with *ab initio* pair interaction energy. As shown below, a new function is required to model the repulsive three-body interactions in bifurcated hydrogen-bonded water trimers, in addition to the usual FQ model as expressed in Eq. (3). Furthermore, we found that the liquid structure and other properties can be dramatically affected by subtle changes in the potential parameters for pair interactions. We attribute this to the significant reduction of pair interactions to simple pair functions. An intelligent mapping of the relevant configuration space of pair interactions is crucial for parametrizing pair interactions.

A. Many-body interactions

Table I suggests that it would require a functional form including all 21 internal nuclear degrees of freedom in order to faithfully model the *ab initio* three-body energies. Complicated and computationally costly potential energy functions and their derivatives may be feasible for molecular dynamics simulations, provided that one can impose a distance cut-off and/or use a neighbor list. In the present paper, we take a different approach.

In molecular dynamics simulations, the most time-consuming part of the calculations is the evaluation of the forces, which involve derivatives of the potential energy function. For the potential energy function defined in Eq. (3),

the force exerted on each atom (denoted by subscripts a, b, and c) is

$$-(\nabla_{\mathbf{r}_{ia}} U)_{\{r_{jb, jb \neq ia}\}} = -(\nabla_{\mathbf{r}_{ia}} U)_{\{r_{jb, jb \neq ia}\}, \{q\}}$$

$$-\sum_{kc} \left(\frac{\partial U}{\partial q_{kc}}\right)_{\{\mathbf{r}\}, \{q_{jb, jb \neq kc}\}}$$

$$\cdot (\nabla_{\mathbf{r}_{ia}} q_{kc})_{\{r_{jb, jb \neq ia}\}}, \tag{45}$$

since the charges $\{q\}$ are not independent variables, but depend on the interatomic distances through the coupled linear equations due to the requirement of electronegativity equalization. The summation on the right-hand side of the above equation appears to introduce an additional complication in the force evaluation. However, this term vanishes because the instantaneous electronegativities (the partial derivative of potential energy with respect to the charges) are equal and because the net molecular charge is conserved (or the charge of the whole system is conserved if intermolecular charge transfer is allowed). Another way of understanding it is to recognize that the charges $\{q\}$ for each configuration $\{\mathbf{r}\}$ are variationally optimized. Thus, the variation of the charges with respect to spatial displacement does not need to be considered in the force evaluation, and the force on each atom is simply

$$-(\nabla_{\mathbf{r}_{ia}} U)_{\{\mathbf{r}_{jb, jb \neq ia}\}} = -(\nabla_{\mathbf{r}_{ia}} U)_{\{\mathbf{r}_{jb, jb \neq ia}\}, \{q\}}. \tag{46}$$

The concept embodied in Eq. (46) has been exploited in computational simulations using polarizable potential models. For example, in the potential models using induced dipoles to account for the many-body effects, the variation of the dipole moments with respect to the spatial displacement also vanishes because the dipole moments are calculated variationally.[48] Thus, the force evaluation for these potential models is as simple as the corresponding nonpolarizable model because the part of the potential function with an explicit dependence on spatial displacement is the same as for the nonpolarizable models. For the empirical FQ potential, the contribution to the force on each atom from the U_{ee} term is simply

$$-(\nabla_{\mathbf{r}_{ia}} U_{ee})_{\{\mathbf{r}_{jb, jb \neq ia}\}} = -\sum_{jb \neq ia} \frac{\partial U_{ee}}{\partial r_{iajb}} \frac{\mathbf{r}_{iajb}}{r_{iajb}}$$

$$= -\sum_{jb \neq ia} q_{ia} q_{jb} \frac{\partial J_{iajb}}{\partial r_{iajb}} \frac{\mathbf{r}_{iajb}}{r_{iajb}}. \tag{47}$$

The force in Eq. (47) appears to be ''pairwise,'' with the many-body couplings incorporated in the charges.

The above discussion is not limited to Coulombic forces. One can also take advantage of Eq. (46) in designing a model for the repulsive three-body interactions. Specifically, the potential function for the short-range nonelectrostatic many-body interaction is taken to be a function of both the coordinates and the charges. The charges, again, are solved by Eq. (6), and thus carry the many-body character. In this approach, the nonelectrostatic many-body interaction would induce a change in the charges, in addition to the energetics. This is consistent with *ab initio* calculations that show qualitatively that the nonelectrostatic and electrostatic many-body interactions differ not only in their effects on the energetics

J. Chem. Phys., Vol. 108, No. 12, 22 March 1998

Liu *et al.* 4749

TABLE II. FQ potential function parameters

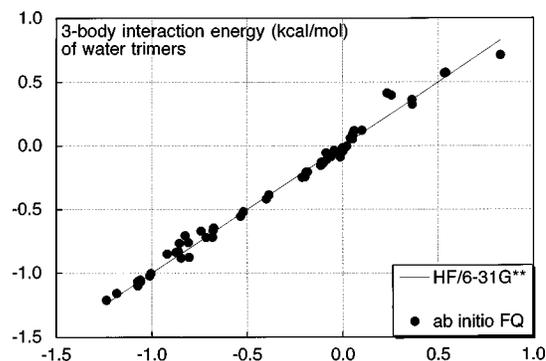|	Empirical[a]	*Ab initio*
Pair parameters		
ϵ_{OO}(kcal/mol)	0.2862	0.3411
$\sigma_{OO}[\text{Å}^{-1}]$	3.159	3.1130
Many-body parameters		
$\widetilde{\chi}_O^0 - \widetilde{\chi}_H^0$[kcal/(mol e)]	68.49	64.97
$\zeta_O(a_0^{-3})$	1.63	1.60
$\zeta_H(a_0^{-3})$	0.90	0.91
s_{OH}[kcal/(mol e)]		13.64
s_{HO}[kcal/(mol e)]		−28.25
s_{HH}[kcal/(mol e)]		−25.36
τ_{HH}		0.0574
τ_{OH}		0.7880
$\sigma_{OO}(\text{Å}^{-1})$		0.7652
$\sigma_{OH}(\text{Å}^{-1})$		1.341
$\sigma_{HH}(\text{Å}^{-1})$		1.475

[a]Reference 11.

FIG. 6. Three-body interaction energies (kcal/mol) of 57 water trimers calculated by the new *ab initio* FQ potential (solid circle) are plotted against *ab initio* results at the Hartree–Fock level. As in Fig. 4, a solid line representing perfect agreement with the HF results is plotted to aid in a visual comparison.

but also the molecular charges. Thus, the aforementioned approach is superior to a many-body function with explicit dependence only on the interatomic distances.

There are many possibilities for the functional form for the nonelectrostatic many-body interactions. In the following, we present one of the many possibilities. In this *ab initio* FQ model, the quantity, χ^0, in Eq. (3) is now a function of interatomic distances, in addition to the constant term, $\widetilde{\chi}^0$,

$$\chi_{i\alpha}^0 = \widetilde{\chi}_\alpha^0 + \sum_\beta \sum_{j \neq i} \sum_\gamma c_{\beta\gamma}^\alpha e^{-(r_{i\beta j\gamma}/\sigma_{\beta\gamma})^2}, \quad (48)$$

where

$$c_{OO}^O = -(s_{OH} + s_{HO}), \quad c_{OO}^H = -s_{HH},$$
$$c_{OH}^O = \tau_{OH}s_{OH}, \quad c_{OH}^H = \tau_{OH}s_{HH},$$
$$c_{HO}^O = \tau_{OH}s_{HO}, \quad c_{HH}^H = \tau_{HH}s_{HH},$$
$$c_{HH}^O = \tau_{HH}(s_{OH} + s_{HO}).$$

The parameters in Eq. (48), $\sigma_{\alpha\beta}$, $s_{\alpha\beta}$, and $\tau_{\alpha\beta}$, along with the $\widetilde{\chi}^0$ and the Slater exponents ζ, are optimized to fit the three-body interaction energy of 57 trimers; under the constraint that the monomer dipole moment being the same as the best *ab initio* estimate.[49] The values of these parameters that produce the optimal fit are given in Table II, and the three-body energies calculated using the *ab initio* FQ parameters are represented by the solid circles in Fig. 6. The maximum absolute error in the three-body energies from the *ab initio* FQ model is 0.15 kcal/mol; and the root-mean-square deviation is 0.05 kcal/mol compared to 0.18 kcal/mol of the empirical FQ potential. For the 43 trimers with attractive three-body energy, the *ab initio* FQ parameters give a value of 1.41 kcal/mol for the total absolute difference between E_{3b}^{FQ} and E_{3b}^{QM}.

B. Pair interactions

To develop the empirical pair potential, we have built an extensive database of water dimer energies at the LMP2 level using the cc-pVTZ(-f) basis set of Dunning and co-workers,[25] which has been previously shown to yield rea-

sonable results for the binding energy of the water dimer. The monomers are constrained to be nearly the equilibrium water monomer geometry, with the largest deviation of 0.04 Å in the OH bond length and 3.15° in the HOH bond angles. The pair potential function in the *ab initio* FQ model is taken to be the usual LJ interaction between oxygens only. Although the possibility of employing more sophisticated and complex pair functions has been explored before, the improvement in the ability to reproduce empirical data is negligible, except when the internal degrees of freedom are included in the molecular dynamics simulations.[50] The present study employs only a rigid water model, and therefore we choose the simple Lennard-Jones function for computational efficiency.

A recent work by Wallqvist and Åstrand shows that subtle changes less than 0.1 kcal/mol in the potential of mean force result in large differences in the liquid density.[51] The *ab initio* data employed in our fitting (or, in general) are not accurate to within 0.1 kcal/mol. This, we believe, limits the advantage of employing sophisticated functional forms for pair interactions. And we have indeed tried to use functions in addition to the Lennard-Jones potential between oxygens, but we did not obtain the kind of overall improvement that would justify the increased cpu time required in simulations with more complex force fields. Therefore, in the present work, we seek correlations between the simulated liquid structures and the *ab initio* water dimer data. In particular, we look for a balance in the representation of long- and short-range interactions.

Caution must be taken, however, to note that simple intermolecular pair functions are not able to reproduce the relative stability of all the conformers of water dimers at a given oxygen−oxygen distance, especially when the monomers are close to each other. To overcome this problem, we include various conformers of water dimers with $R_{OO} > 4$ Å, but only those that resemble the equilibrium gas-phase dimer conformation (i.e., nearly linear hydrogen bonded) are included in the fitting for R_{OO} less than 3 Å. In summary, the dimer dataset includes totally 94 dimers, with the oxygen−oxygen distance ranging from 2.66 to 3.00 Å (28 dimers) and from 4 to 7 Å (66 dimers); and their binding energies are all less than 1.4 kcal/mol. This energy cutoff is chosen because the

TABLE III. Static and dynamics properties of water.

	Hartree–Fock cc-pVTZ(-f)	FQ models		Experiment[b]
		Empirical[a]	*Ab initio*	
Gas-phase properties				
Dipole moment (Debye)	2.0299	1.85	1.85	1.85
α_{zz} (Å^3)	0.9969	0.82	0.86	1.468±0.003
α_{yy} (Å^3)	1.184	2.55	2.50	1.528±0.013
α_{xx} (Å^3)	0.7487	0.0	0.0	1.415±0.013
Dimer energy (kcal/mol)		−4.51	−4.60	−5.4±0.7
Dimer O–O length (Å)		2.92	2.91	2.98
Liquid-phase properties				
Energy (kcal/mol)		−9.9	−10.1	−9.9
Dipole moment (Debye)		2.62	2.68	
τ_D (ps)		8±2	8±3	8.27±0.02
Dielectric constant ϵ_0		79±8	79±16	78
Dielectric constant ϵ_∞		1.592±0.003	1.593±0.005	1.79
Diffusion constant (10^{-9} m^2/s)		1.9±0.1	1.9±0.3	2.30
τ_{NMR} (ps)		2.1±0.1	2.3±0.2	2.1

[a]Reference 11.
[b]The source of experimental data may be found in Ref. 11.

binding energy corresponding to the equilibrium dimer configuration is about −5 kcal/mol, making the very repulsive configurations rare events during simulations. The distribution of the binding energies of the water dimers in the fitting data is as follows, 19 of those less than −4 kcal/mol, 7 of those between −4 and −3 kcal/mol, 1 between −3 and −2 kcal/mol, 5 between −2 and −1 kcal/mol, 40 between −1 and 0 kcal/mol, 21 between 0 and 1 kcal/mol, and 1 above 1 kcal/mol. The Lennard-Jones parameters for the oxygen–oxygen interactions are varied to minimize the deviation from the *ab initio* results. The values for these parameters that give an optimal fit are tabulated in Table II. This set of parameters combined with the functions discussed in Sec. V A gives a maximum deviation of 0.5 kcal/mol from the *ab initio* binding energy and the sum of the absolute values of the deviation is 8.44 kcal/mol.

C. Predicted gas-phase properties

A comparison of equilibrium water dimer geometry and binding energy of the *ab initio* FQ model and the empirical data is presented Table III, along with the empirical FQ potential. The agreement with the experiment is reasonable, and it is superior than most of the other empirical water potential models.[1] We also note that the LMP2/cc-pVTZ(-f) estimate for the water dimer binding energy may be a few tenths of a kcal/mol more positive than higher level quantum chemical calculations, with large basis sets that currently yield results on the order of 5.0 kcal/mol.[52]

Since our focus here is to design a polarizable force field for water, it is interesting to examine the dipole polarizabilities of the new *ab initio* FQ model. A comparison with the empirical FQ potential and experiment is presented in Table III. We define the xyz axes such that the C_2 axis of the water molecule is along the z axis, and the molecule is in the yz plane. Since all molecular charge sites are in the molecular plane, the out-of-plane dipole polarizability α_{zz} is zero for both the *ab initio* and empirical FQ potentials, while the

empirical dipole polarizability is nearly isotropic. The in-plane polarizabilities from both the empirical and the *ab initio* FQ models are very different from the experimental estimate. As discussed in Sec. IV, a three-site FQ model is incapable of modeling the polarization response to uniform and nonuniform fields simultaneously, and distinct sets of potential parameters are required to fit the polarization response in the presence of a nonuniform field and to fit three-body energies of water trimers. It is therefore not surprising that the molecular polarizabilities of our new *ab initio* FQ model do not agree with empirical values. How this affects modeling a truely transferable force field is further discussed in Sec. VI.

D. Predicted liquid-phase properties

The predicted static and dynamics properties of liquid water at 298 K and $\rho = 1$ g/cm^3 are presented in Table III, and they are compared to the empirical FQ potential and experiment. The error bars represent two standard deviations.
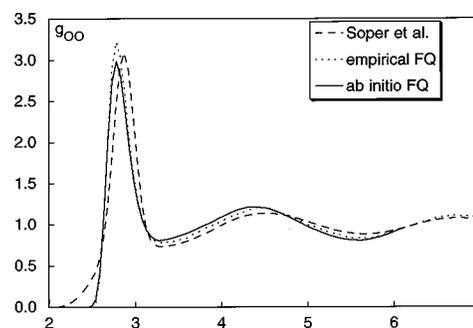

FIG. 7. Oxygen–oxygen radial distribution function for the *ab initio* FQ (solid line) and the empirical FQ (dotted line) potentials, compared to the neutron diffraction results of Soper and Phillips (dashed line).
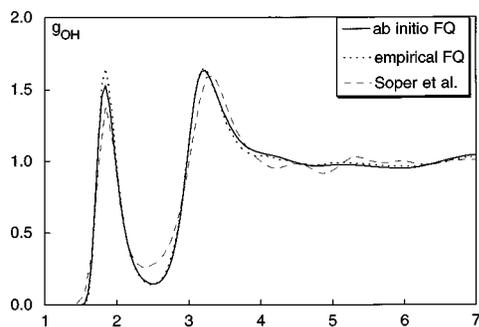
FIG. 8. Oxygen–hydrogen radial distribution function for the *ab initio* FQ (solid line) and the empirical FQ (dotted line) potentials, compared to the neutron diffraction results of Soper and Phillips (dashed line).

The internal energy predicted by the *ab initio* FQ model agrees with the experiment estimate reasonably well.

The radial distribution functions (g_{OO}, g_{OH}, and g_{HH}) of intermolecular oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen distance are presented in Figs. 7–9. The prediction of the *ab initio* FQ potential is represented by the solid lines, and the results of the empirical FQ potential and the experiment by Soper and Philips[53] are represented by the dotted and the dashed lines, respectively. For comparison, we also plotted the *ab initio* FQ pair correlation functions in Fig. 10 with two sets of neutron diffraction results: one by Soper and Philips[53] and the other one by Soper and Turner.[54] As seen in Fig. 10, the peak positions obtained from the two neutron diffraction measurements show much less uncertainty than the peak height. The liquid structure predicted by the *ab initio* FQ model agrees very well with the empirical FQ potential, and they both are in good agreement with the experiment. Nevertheless, the position of the first peak of both FQ models exhibits a notable deviation from the experiment. In fact, it is well known that all the cited force fields predict a shorter most probable first shell O–O distance than the experiment by approximately 0.1 Å. The first peak position of g_{OO} predicted by the FQ potentials is in closer agreement with the experiment than the nonpolarizable water models. Two factors reduce the significance of this difference. First, the experimental diffraction measurements must be Fourier inverted and separated into O–O, O–H, and H–H components. Over the years Soper *et al.* have reported different experimental results. Although the major differences
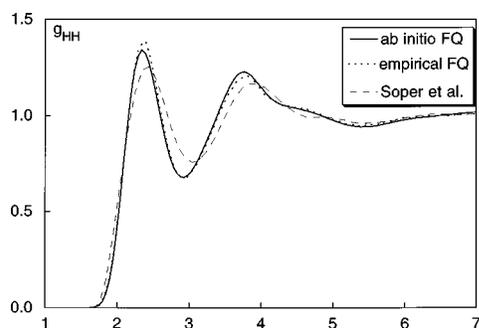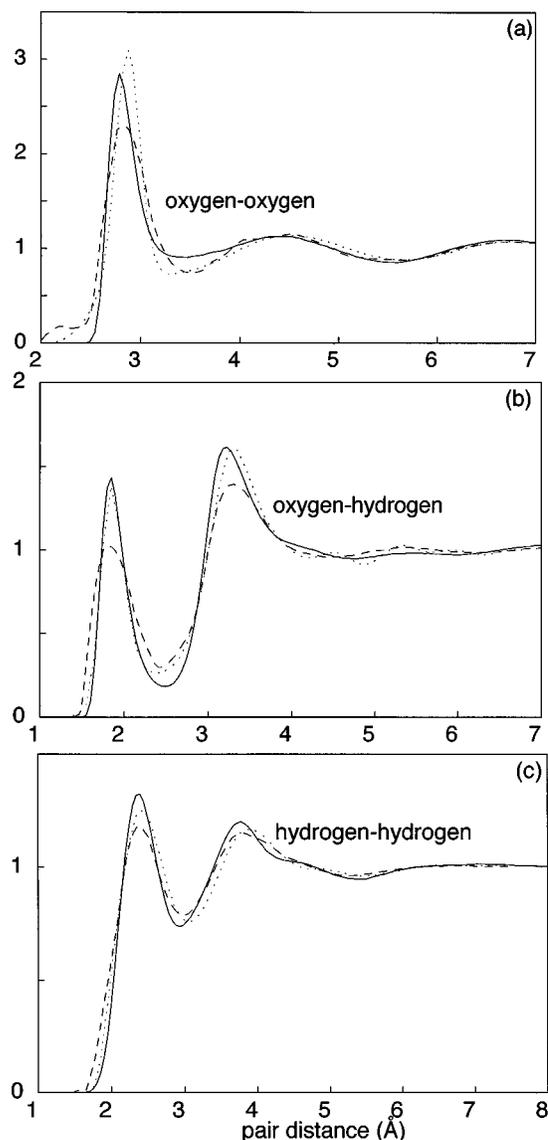

FIG. 10. Radial distribution functions for the *ab initio* FQ model (solid line) compared to two sets of neutron diffraction results, by Soper and Phillips (dashed line)[53] and by Soper and Turner.[54] Panels (a), (b), and (c) are oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen pair correlation functions, respectively.

have been in the peak heights, this changing experimental target does not engender confidence that we have seen the final result. Second, it should be noted that the force field does a very good job of getting the dimer O–O distance correctly. Last, classical simulations were done. Were quantum path-integral simulations performed the peak positions in the liquid would probably shift out and the peaks would get somewhat broader because of quantum zero-point fluctuations. It is encouraging to see an equally good prediction from the *ab initio* FQ model, even though the structural data are not included in the fitting of this water potential model.

The predicted liquid-state dipole moment from the *ab initio* FQ model is in good agreement with the empirical FQ potential. In Fig. 11 we compare the distribution of liquid-state water dipole moments of the *ab initio* and empirical FQ potentials. The maxima of the two distribution curves are very close to each other, whereas the distribution of the *ab*

FIG. 9. Hydrogen–hydrogen radial distribution function for the *ab initio* FQ (solid line) and the empirical FQ (dotted line) potentials, compared to the neutron diffraction results of Soper and Phillips (dashed line).
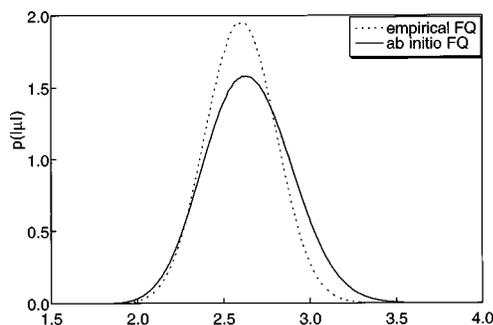
FIG. 11. A distribution of the absolute value of the liquid-phase dipole moments (in Debye) of the *ab initio* FQ (solid line) and the empirical FQ (dotted line) potentials.

FIG. 13. The real part of the frequency-dependent dielectric constant for the *ab initio* FQ model (solid lines), compared to the empirical FQ potential (dashed line) and experiment (dotted lines).

initio FQ water dipole exhibits a wider width, and, in particular, a more pronounced tail distribution of large dipole moments. Thus, it is interesting to compare the predicted dielectric constants of the two FQ models, which give very close magnitudes for the averaged liquid-state dipole moments, but differ in the dipole moment distribution. The predicted optical and static dielectric constants from the *ab initio* FQ model are compared to those of the empirical FQ potential and experiment in Table III, and they are in very good agreement with one another and with the experiment. Another dynamical property related to the liquid-state dipole moment is the Debye relaxation time, τ_D, which can be obtained from the time-correlation function (ϕ) of the averaged liquid-state dipole moments shown in Fig. 12. In the present case, τ_D is estimated by fitting the long-time tail of ϕ to the function $A \exp(-t/\tau_D)$. We find very good agreement between the *ab initio* and empirical FQ potentials, and also very good agreement with the experiment. The agreement between the two FQ models is consistent with the close agreement of the liquid-state dipole moments predicted by the two models.

By taking a Laplace transformation of the time-correlation function of the averaged liquid-state dipole moments, one obtains the frequency-dependent dielectric constants. In Figs. 13 and 14 the real and the imaginary parts of the frequency-dependent dielectric constants of the *ab initio* and empirical FQ potentials are compared, along with the experimental values. Again, we find very good agreement

between the two FQ models and also with the experiment. Figure 14 corresponds to the dielectric saturation spectrum. Its intensity decreases when the molecules fail to respond to the electric field of the electromagnetic wave. For frequency less than 40 THz, molecules respond to the electric field by intermolecular relative translation and libration motion. For comparison, we calculate the normal mode frequencies of the water dimer, trimer, and tetramer at their minimum energy configurations. Since both the empirical and *ab initio* FQ potentials are for rigid water molecules, one must project out the components due to intramolecular degrees of freedom from the Hessian matrix. To achieve this, we first construct a density matrix from the unit vectors corresponding to the translation and rotation of each water molecule, and then project out the density corresponding to the overall translation and rotation of the complexes. This density matrix is then applied to the Hessian matrix to exclude the components due to intramolecular motions. The intermolecular modes of the water dimer are all between 20 and 30 THz, with an exception of the lowest frequency being 6 THz. The former corresponds to librations and the latter corresponds to the breaking of the hydrogen bond. As more water molecules are included in the complex, collective modes are formed and the distribution of the frequencies extends into the microwave region. The experimental observation of the libration motion is around 15 THz (see Fig. 14). The prediction of
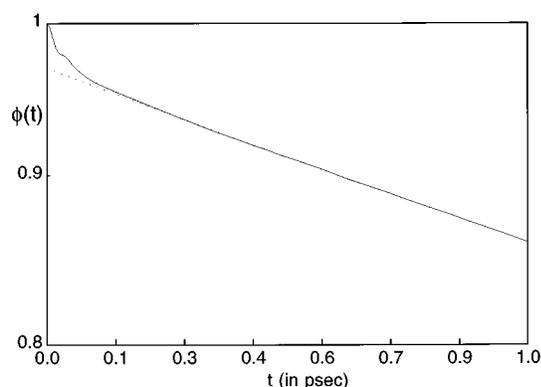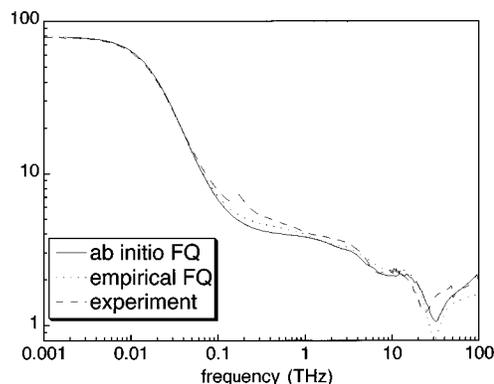

FIG. 12. The time autocorrelation function of the liquid-phase dipole for the *ab initio* FQ (solid line) and the empirical FQ (dotted line) potentials.
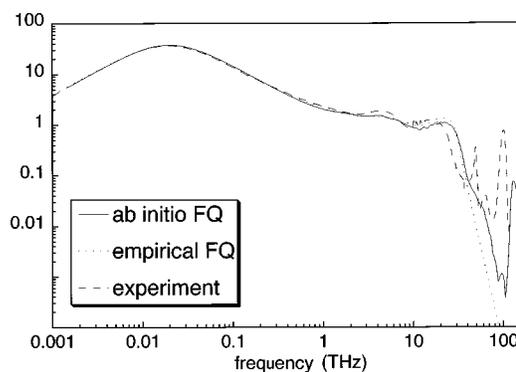

FIG. 14. The imaginary part of the frequency-dependent dielectric constant for the *ab initio* FQ model (solid lines), compared to the empirical FQ potential (dashed line) and experiment (dotted lines).

both the empirical and the *ab initio* FQ potentials is blue-shifted. This may be an artifact of the rigid molecule model, and we expect these librational modes to be softened when coupling to intramolecular modes is permitted. In fact, we note that the simulations using a flexible water model gives perfect agreement with the experiment in terms of the peak position of the librational modes.[55]

The highest normal mode frequency of the intermolecular modes increases slightly as the number of water molecules increases. The sharp peak around 130 THz exhibited by both FQ models is not a result of motion of water molecules. This region of the spectra corresponds to intramolecular stretching and bending modes, which is absent in both FQ models. Rather, this peak is a result of charge oscillation about the ''adiabatic'' values [i.e., values satisfy Eq. (6)]. Equation (33) may be rewritten as

$$\mathscr{L} = \sum_{i=1}^{N_{mole}} \sum_{\mu=1}^{N_{atom}} \frac{1}{2} M_\mu \dot{r}_{i\mu}^2 + \sum_{i=1} \sum_{j>i} \sum_{\mu}^{N_{atom}} \sum_{\nu} V(r_{i\mu j\nu})$$
$$- \sum_{\gamma=1}^{N_{bond}} \Gamma_\gamma G_\gamma(\{\mathbf{r}\}) + \frac{m_q}{2} \dot{\mathbf{q}}'^\dagger \dot{\mathbf{q}}' + \mathbf{q}'^\dagger \mathbf{CX}$$
$$+ \frac{1}{2} \mathbf{q}'^\dagger \mathbf{CJC}^\dagger \mathbf{q}', \qquad (49)$$

where **C** transforms the coordinates **q** to **q'** as, for example, in Eq. (9). Note that here **C** is defined such that the kinetic energy of the charges retains the diagonal form in the new mass-scaled coordinates **q'**, in addition to eliminating the redundant coordinates due to the constraints [Eqs. (7) and (8)]. The term $\mathbf{CJC}^\dagger$ in the above equation is the ''force matrix,'' and it may be diagonalized to obtain the ''normal mode frequencies'' of the charges. Take a monomer as an example, one obtains 147 and 251 THz for the charge oscillations. These values support our explanation of the *ab initio* results in the high-frequency region of Fig. 14.

Also presented in Table III are the diffusion constant and the NMR relaxation time, τ_{NMR}. The former is estimated using the Einstein relation, and the latter is calculated from the long-range component of the second-order rotational time constants corresponding to the rotation about the axis connecting the hydrogen atoms. Both quantities predicted by the *ab initio* model are comparable to that of the empirical FQ potential, and are superior to typical values obtained from empirical nonpolarizable pair potentials. The vapor pressure of the *ab initio* FQ water model at 298 K, however, is about −0.6 kbar (compared to the experimental value of 0.001 kbar). This result is less satisfactory. Nevertheless, we note that the error introduced in the density due to this discrepancy is approximately 2.7%, raising it from 0.997 to 1.024 g/cm³ (based on liquid water compressibility). Therefore the *ab initio* FQ water model should also give reliable water properties in a constant pressure simulation. Very small adjustments (within the accuracy of the present level of *ab initio* calculations) can be made to give more accurate vapor pressures. When a more sophisticated functional form is used for the pair part of the potential, allowing a better fit to the full *ab initio* dataset, we expect these results to improve.

VI. DISCUSSIONS AND CONCLUSIONS

Several conclusions may be drawn from this paper. First, we have demonstrated in the case of water that it is possible to obtain an *ab initio* polarizable force field as accurate as the most successful empirical polarizable potentials. This is achieved by first extracting information of many-body interactions from studies of water trimers with various relative orientation and intermolecular distances. The potential parameters pertinent to many-body interactions, the FQ parameters in the present case, are then adjusted to give optimal agreement between the FQ model and the *ab initio* calculations. A new function is introduced to model the repulsive three-body interactions exhibited in the bifurcated hydrogen-bonded water trimers, while at the same time keeping the intermolecular forces as simple functions of interatomic distances. Parameters that affect only the pair interactions are determined last. We found that subtle changes in the pair potentials can cause dramatic changes in the liquid structure and other properties such as pressure. We attribute this to the significant reduction of complicated pair interactions to the simple pair functions entailed by large-scale molecular dynamics simulations. The pair part of our *ab initio* water potential is parametrized to long-range pair interactions (the distance between the oxygens beyond 4 Å), and the short-range interactions between hydrogen-bonded pairs of water (the distance between the oxygens less than 3 Å). This parametrization scheme enables us to obtain very good agreement in the static and dynamics properties with experiment.

The limitations due to the simplicity of the potential functions not only are observed in the modeling of pair interactions, but also appear to affect many-body electrostatic polarization. In Sec. IV, the potential parameters responsible for electrostatic polarization (i.e., the Slater orbital exponents ζ_O and ζ_H) are varied to fit the *ab initio* polarization response of a water molecule in the presence of a nonuniform external electric field. We found that the fitted potential parameters lead to molecular polarizabilities that are very different from the *ab initio* results at exactly the same level of theory and using exactly the same basis functions. That is, distinct sets of molecular polarizabilities (and therefore distinct sets of potential parameters) are required to represent the polarization response to uniform and nonuniform external electric fields. This is due to the fact that the number of potential function parameters in our water model is not sufficient for modeling the response to both types of external fields simultaneously. This limitation is clearly shown in Eq. (15), which is rewritten below:

$$\mathbf{A}\Delta\mathbf{Q} = -\mathbf{v}', \qquad (50)$$

where **A** denotes $\mathbf{CJC}^\dagger$, and **v'** is **Cv**. The polarization response, $\Delta\mathbf{Q}$ (i.e., the change in charge distribution due to the external field), depends only upon the values of the potential at the charge sites in the model. Thus, if there are two different external potentials, $\mathbf{v}_1$ and $\mathbf{v}_2$, which have identical values at the charge sites (a condition that is trivially satisfied, for example, by taking linear combinations of field sources and solving for the coefficients), but induce signifi-

cantly different polarization responses in an accurate quantum chemical calculations, the model is fundamentally incapable of reproducing any such differences.

To what extent can the FQ model properly resolve variations in the polarization response arising from spatial inhomogeneity? This is most easily understood via an eigenvalue/eigenvector decomposition of Eq. (50). $\mathbf{A}$, a real symmetric matrix $[(N_{mole}N_{site}-N_{mole})$ by $(N_{mole}N_{site}-N_{mole})$ for constraint Eq. (7)], has a set of eigenvectors ψ_j that satisfy the equation

$$\mathbf{A}\psi_j = J_j\psi_j, \tag{51}$$

where the J_j are the corresponding (real) eigenvalues. The solution for $\Delta\mathbf{Q}$ can then be written as

$$\Delta\mathbf{Q} = \sum a_j J_j^{-1}\psi_j, \tag{52}$$

where $a_j = \langle\psi_j|\mathbf{v}\rangle$ is the projection of the jth eigenvector on the applied field vector. The magnitude of the polarization response to an applied field can thus be decomposed into different amplitudes J_j^{-1}, each of which is associated with a particular spatial pattern of the field as it is resolved by the charge sites.

Thus, in our FQ model for water, there are three charge sites, and hence, for any value of the FQ parameters, exactly two spatially resolved polarization responses. If the accurate quantum chemical polarization response is fit well by two eigenvalues (it can be decomposed in exactly the same fashion, except in this case there are, in principle, an infinite number of modes), the current model can be made to reproduce this via parameter adjustment. On the other hand, if there are a larger number of important modes, the fit can encompass only a limited subset of these, and will invariable make errors in others. This explains why it is that different parameter values result from fitting to different chemical datasets.

However, the model can be systematically improved by adding charge sites. The key to limiting the number of charge sites, and hence retaining computational efficiency, is to determine what regions of physical space are crucial to resolving spatially inhomogeneous variation—for example, one possibility is that a better representation of the oxygen lone pairs is necessary, in which case charge sites could be placed in these locations. This information can be readily obtained via the appropriate quantum chemical experiments, and will be the subject of a future publication.

The conclusion of this analysis is therefore that, as long as the linear response regime holds, the FQ model, *in principle*, is capable of providing an arbitrarily accurate reproduction of the quantum chemical behavior, simple by increasing the number of charge sites until the entire spectrum of inhomogeneous behavior is properly reproduced. And linear response appears to be perfectly adequate for the magnitudes of the perturbations that we have examined, which are typical of molecular systems. Of course, it is possible that there are specific molecules or environments in which the linear response is not completely reliable; however, the linear response is still likely to be the dominant term and hence a significant improvement over fixed charge models, even in

these cases. What remains to be determined is whether the degradation of the computational efficiency as a consequence of adding enough charge sites to converge the response function will be substantial. In practice, it is likely that compromises will have to be reached, depending upon the available computer resources and the goals of the simulation.

In Sec. IV we also try to establish correlations between the three-body energy and the molecular polarizabilities. Our results show that potential function parameters that give a reasonable representation of the three-body energies of water trimers from quantum chemical calculations could lead to a wide range of values for α_{yy}, the y component molecular polarizability (i.e., along the HH axis), and thus suggest little correlations between the three-body energies and α_{yy}. From simulations of pure water, the averaged induced dipole along the HH axis is nearly zero, as expected, if the solvent response is regarded as a dielectric continuum reaction field. In this case, a solute water molecule would induce a net solvent reaction field along the C_2 molecular axis, but not along the y direction. It is difficult, however, to extend such an explanation appropriate for bulk systems to small molecular clusters. Moreover, one would expect the y component of the induced dipole should have nonvanishing effects on the instantaneous dynamics of individual water molecules in the bulk, and thus affect the relaxation dynamics observed in simulations. Therefore, it is important to understand the reasons behind the success of our *ab initio* FQ model presented in Sec. V. In other words, is it due to the compensation of out-of-plane polarization by the relatively large value of α_{yy}, or that only the polarization along the C_2 axis matters? A great deal of additional work will be required to answer this question. Furthermore, the differences between the two approaches for modeling many-body interactions, as manifestations of monomer properties or as captured in molecular clusters, are of great importance to establishing a general and efficient procedure for constructing accurate *ab initio* force fields. These issues are currently under investigation.

ACKNOWLEDGMENTS

[1] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[2] See the entire issue of number 7 of Chem. Rev. **94**, 1721 (1994).
[3] K. Morokuma, J. Chem. Phys. **55**, 1236 (1971).
[4] H. Umeyama and K. Morokuma, J. Am. Chem. Soc. **99**, 1316 (1977).
[5] A. Wallqvist, P. Ahlström, and G. Karlström, J. Phys. Chem. **94**, 1649 (1989).
[6] P.-O. Åstrand, P. Linse, and G. Karlström, Chem. Phys. **191**, 195 (1995).
[7] J.-C. Soetens and C. Millot, Chem. Phys. Lett. **235**, 22 (1995).

[8] G. Chalasiński and M. M. Szczęśniak, Chem. Rev. **94**, 1723 (1994).

[9] C. Millot and A. J. Stone, Mol. Phys. **77**, 439 (1992).

[10] E. Clementi and G. Corongiu, ACS Symp. Ser. **568**, 95 (1994).

[11] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).

[12] R. T. Sanderson, Science **114**, 670 (1951).

[13] H. S. Frank and W. Y. Wen, Proc. R. Soc. London, Ser. A **247**, 481 (1958).

[14] F. H. Stillinger, Science **209**, 4455 (1980).

[15] R. P. Iczkowski and J. L. Margrave, J. Am. Chem. Soc. **83**, 3547 (1961).

[16] R. S. Mulliken, J. Chem. Phys. **2**, 782 (1934).

[17] R. G. Parr and R. G. Pearson, J. Am. Chem. Soc. **105**, 7512 (1983).

[18] R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, J. Chem. Phys. **68**, 3801 (1978).

[19] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

[20] A. K. Rappe and W. A. Goddard, J. Phys. Chem. **95**, 3358 (1991).

[21] M. N. Ringnalda, J.-M. Langlois, B. H. Greeley, R. B. Murphy, T. V. Russo, C. Cortis, R. P. Muller, B. Marten, R. E. Donnelly, D. T. Mainz, J. R. Wright, W. T. Pollard, Y. Cao, Y. Won, G. H. Miller, W. A. Goddard, III, and R. A. Friesner, PS-GVB v2.2, Schrodinger, Inc., 1995.

[22] S. Sæbø and P. Pulay, Theor. Chim. Acta **69**, 357 (1986).

[23] S. Sæbø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).

[24] R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Chem. Phys. **103**, 1481 (1995).

[25] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[26] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[27] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).

[28] D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., unpublished.

[29] For a recent review, see F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).

[30] K. Kim and R. A. Friesner (private communication).

[31] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).

[32] M. Parrinello and A. Rahman, Phys. Rev. Lett. **45**, 1196 (1980).

[33] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[34] S. Nosé, Mol. Phys. **52**, 225 (1984).

[35] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Baldus, Jr., Phys. Rev. Lett. **49**, 1691 (1982).

[36] F. J. Vesely, J. Comput. Phys. **24**, 364 (1977).

[37] P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, Nature (London) **282**, 459 (1979).

[38] P. Ahlström, A. Wallqvist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).

[39] E. Clementi, in *Modern Techniques in Computational Chemistry* (Escom, Leiden, 1989).

[40] G. Ruocco and M. Sampoli, Mol. Phys. **82**, 875 (1994).

[41] D. N. Bernardo, Y. Ding, K. Krogh-Jespersen, and R. M. Levy, J. Comput. Chem. **16**, 1141 (1995).

[42] L. N. G. Filon, Proc. R. Soc. Edinburgh Sect. A **49**, 38 (1928).

[43] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).

[44] F. H. Stillinger, Adv. Chem. Phys. **31**, 1 (1977).

[45] G. Chalasiński, M. M. Szczęśniak, C. Cieplak, and S. Scheiner, J. Chem. Phys. **94**, 2873 (1991).

[46] Prior quantum chemical evaluations of molecular polarizability show that *ab initio* results agree with empirical data to within 16% when the diffuse functions are included in the basis set, and the agreement is further improved to within 3% when the calculations are carried out at the MPC level and beyond.[47] Our own analysis shows that the molecular polarizabilities, α_{xx}^{QM}, α_{yy}^{QM}, and α_{zz}^{QM} are 0.7870, 1.2206, and 1.0370 Å^3, respectively, at the MP2 level using the cc-pVTZ basis set. These results are slightly increased from the values calculated at the Hartree–Fock level, but are still substantially smaller than the experimental values, which range from 1.4 to 1.5 Å^3.

[47] C. Voisin, A. Cartier, and J. Rivail, J. Phys. Chem. **96**, 7966 (1992).

[48] F. H. Stillinger, J. Chem. Phys. **71**, 1647 (1979).

[49] D. Feller, J. Chem. Phys. **98**, 7059 (1993).

[50] See Refs. 1, 56 and 57 for a comparison.

[51] A. Wallqvist and P.-O. Åstrand, J. Chem. Phys. **102**, 6559 (1995).

[52] For a detailed discussion, see S. Scheiner, Annu. Rev. Phys. Chem. **45**, 23 (1994).

[53] A. K. Soper and M. G. Phillips, Chem. Phys. **107**, 47 (1986).

[54] A. K. Soper and J. Turner, Int. J. Mod. Phys. B **7**, 3049 (1993).

[55] J. Martí, E. Guàrdia, and J. A. Padró, J. Chem. Phys. **101**, 10 883 (1994).

[56] K. Watanabe and M. L. Klein, Chem. Phys. **131**, 157 (1989).

[57] J. R. Reimers and R. O. W. M. L. Klein, Chem. Phys. **64**, 95 (1982).