Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations

Jiali Gao

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260

Received: September 13, 1996; In Final Form: October 28, 1996[⊗]

A new method for the treatment of many-body polarization effects in fluid systems is presented, making use of hybrid QM/MM techniques and the semiempirical wave function. In this approach, the electronic structure of each solvent molecule is represented by an antisymmetric determinant wave function, which can be determined by fully converged variational QM calculations at each step of the classical trajectory or can be treated as dynamic variables along with the nuclear coordinates in molecular dynamics simulations. In determining the molecular wave functions, a hybrid QM/MM method is used. This molecular orbital derived empirical potential for liquid simulations (MODEL) is analogous to the fluctuating charge model introduced by Berne and co-workers; however, the MODEL potential is a quantum chemical model as opposed to a purely empirical charge equilibration scheme employed in the FC model. The method may also be extended for the treatment of large molecular systems. The capability of the MODEL potential for describing bimolecular interactions was illustrated for bimolecular complexes. Reasonable agreement with ab initio results suggests that the MODEL potential may be parametrized for liquid simulations.

1. Introduction

Statistical mechanical Monte Carlo and molecular dynamics simulations have become an essential tool for the study of ensemble properties of molecular systems.[1,2] These calculations are largely carried out with the use of empirical potential functions or molecular mechanics (MM) force fields.[3] Thus, the energy of the system and forces upon the atoms can be rapidly evaluated. In most applications, pairwise effective potentials are used, and the electronic structure of the molecule is approximated by partial atomic charges fixed on well-defined sites in the molecular frame. The instantaneous many-body polarization effect in a fluid system is not explicitly treated, and the charges do not change in response to variations in electrostatic fields due to movement of atoms during the simulation. Instead, the polarization effect is modeled in an average fashion during the parametrization process. Although MM effective potentials can provide valuable insight into the equilibrium properties such as the free energy of solvation of a solute in a solution, the form of the empirical potential functions is not appropriate for describing chemical processes involving bond formation and breaking. Furthermore, the lack of the instantaneous polarization effect in pairwise potentials is a serious limitation, especially in the study of molecular recognition and protein−DNA interactions in aqueous solution.

Progress has been made toward the treatment of many-body polarization effects in condensed phases, which go beyond pair potentials.[4−16] However, just as complicated as the phenomenon itself, there are numerous ways to calculate the polarization energy. One of the most widely used approaches is based on the classical expression[17]

$$E_{\text{pol}} = -\frac{1}{2}\sum_{i=1}^{N}\mu_i \cdot \mathbf{E}_i^\circ \qquad (1)$$

where $\mathbf{E}_i^\circ$ is the electric field at the *ith* atomic site resulting from the permanent charge distribution of the system, and μ_i is the converged induced dipole moment. A shortcoming of this method is that atomic polarizabilities $\{\alpha_i\}$ are needed to

determine μ_i, whereas $\{\alpha_i\}$ cannot be uniquely obtained from experimental molecular polarizability.[18] In addition, higher order multipole moments are not explicitly treated in eq 1, and a significant increase in computing resources is necessary to achieve convergence of the induced dipoles. Recently, Berne and co-workers introduced a fluctuating charge (FC) model,[15] which was derived on the basis of the principle of electronegativity equalization—a concept first used in deriving atomic charges and has been extensively investigated by Parr and others using density functional theory.[19−24] The concept was also employed in an MM force field by Rappé and Goddard.[23] Since the principle of electronegativity equalization includes key characteristics of electronic structure, the FC model, although an empirical method, represents a leap forward in the development of polarizable potentials. Along the same line, York and Yang described a procedure based on chemical potential equalization from the perspective of density functional theory for molecular simulations.[24] Nevertheless, it should be emphasized that the polarization effect can only be adequately treated by quantum chemical methods because the molecular charge polarization is a property of the electronic structure in nature.[25]

Indeed, one of the most exciting developments, exemplified by the Car−Parrinello method, is the advent of ab initio molecular dynamics, in which interatomic forces are determined directly from the electronic structure, rather than from empirical potentials.[26] In this approach, in the form of either density functional theory (DFT) or molecular orbital (MO) representation,[26−28] the "orbital" coefficients are treated as a set of fictitious dynamic variables that are propagated adiabatically with respect to the nuclei. Consequently, at each time step during the molecular dynamics simulation, the electronic degrees of freedom are annealed to the Born−Oppenheimer potential energy surface. Although this would be the ultimate goal and indeed the method has been applied to small, periodic systems,[29] it would be difficult in the near future to model the full dynamics of an extended molecular system such as an aqueous solution or a protein because of its computational expense. For example, in a recent estimate of computational bottlenecks to treating large molecules, Head-Gordon pointed out that a Hartree−Fock (HF) or DFT calculation of a protein of about 10 000 atoms

---

would only be remotely feasible with the assumption of a 100-fold increase in computing resources and a true linear scaling in computational complexity with molecular size.[30] A single electronic structure calculation can undoubtedly give valuable insight into the interatomic interactions; however, dynamic and thermodynamic properties of condensed phase systems including proteins must be determined by statistical mechanical Monte Carlo and molecular dynamics methods so that the computed results can be compared with experiment. This leads to an increase of at least 10^6-fold in computing resources because electronic structure calculations must be repeated for each microscopic state sampled in the simulation. Here, we are forced to encounter a "statistical-mechanical computational bottleneck" that one must overcome in ab initio molecular dynamics simulations.

Clearly, the only possibility that ab initio or Hartree−Fock molecular dynamics calculations may be applicable to biological systems is to introduce approximations in the method. Two directions may be envisioned: the use of semiempirical models and the development of novel techniques for treating large molecular systems. In this paper, we describe a molecular orbital derived empirical potential for liquid simulations (MODEL), making use of semiempirical quantum chemical models and hybrid quantum mechanical and molecular mechanical (QM/MM) techniques.[31] The MODEL potential is derived on the basis of quantum chemical representation of individual molecules in a liquid, while the interactions of each monomer with the surrounding molecules are determined by a hybrid QM/MM approach. The latter method has been developed for the study of solute−solvent interactions.[31−35] This paper discusses the algorithms of the MODEL potential, in which the electronic structure and charges are responsive to environmental changes, and results for bimolecular interactions. A second paper describes a new simulation method and parametrization for Monte Carlo and molecular dynamics calculations of pure liquids, and a third sets the theory for treating large molecules and its applicability to protein systems.

In what follows we first outline the theory behind semiempirical methods and the hybrid QM/MM techniques and discuss the approximations that have been made in the MODEL potential and empirical parameters that have been introduced to compensate for these approximations. Section 3 discusses the implementation of the method, and section 4 gives some results for bimolecular complexes. The paper concludes with a summary of the main points of the paper.

2. Theory

We consider a system of N solvent molecules in a cubic box with periodic boundary conditions. Each molecule contains $2M$ electrons and A nuclei (atoms) with effective nuclear charges $\{Z_\alpha\}$. The restricted HF wave function, Ψ_a, which is assumed to be a closed-shell system in the present discussion, is a single antisymmetric determinant of M doubly occupied orthonormal molecular orbitals (MOs), $\{\phi_i\}$:[36]

$$\Psi = |\phi_1(1)\,\alpha(1)\,\phi_1(2)\,\beta(2)...\phi_M(2M)\,\beta(2M)\rangle \quad (2)$$

where α and β are the electron-spin eigenfunctions and the MOs are linear combinations of N_b atomic basis functions, $\{\chi_\mu\}$.

$$\phi_i = \sum_\mu^{N_b} c_{\mu i}\chi_\mu \quad (3)$$

In the neglect of diatomic differential overlap (NDDO) approximation[37] that is made by the MNDO,[38] AM1,[39] and PM3[40]

methods, the orthonormal constraint equations that the MO coefficients must obey are given as follows:

$$\Lambda_{ij} = \sum_\mu^{N_b} c_{\mu i}c_{\mu j} - \delta_{ij} = 0 \quad (4)$$

Note that the constraint conditions are independent of the overlap matrix and thus the nuclear coordinates under the NDDO approximation because the overlap integrals between basis functions on different atoms are assumed to be zero, and the overlap integrals between different basis functions on the same atom are zero by symmetry. This leads to a convenient separation of the electronic variables from nuclear coordinates for molecular dynamics simulations.[28] In the following, indices for nuclei are given in lower case Roman letters a, b, ..., molecular orbitals in lower case Roman letters i, j, ..., and atomic orbitals in lower case Greek letters μ, ν, ... The wave function for the liquid system, Φ, is approximated by a Hartree product of the individual molecular wave functions:

$$\Phi = \Psi_1\Psi_2...\Psi_N \quad (5)$$

Thus, with this approximation, the electronic structure of each solvent molecule is influenced by the average field due to all other solvent molecules, and the correlation effects between solvent molecules are not explicitly treated. For molecules that are separated by relatively long distances, the average electrostatic potential may be represented by a multipole expansion and included in quantum chemical calculations through an hybrid QM/MM approach. However, at shorter distances, for example, close to chemical bonding range, electron correlation effects become important, and an appropriate treatment of the overlap between electron clouds due to the Pauli exclusion principle would also be necessary. With the use of a multipole expansion method to describe electrostatic interactions, Dewar and Thiel formulated an elegant procedure to deal with the electron correlation effects through integral evaluations for use in semiempirical methods.[41] This approach has been largely responsible for the success of hybrid QM/MM models in previous studies.[42] If we limit the multipole expansion only to the monopole term, i.e., point charges, the effective Hamiltonian of the system can be written as follows:

$$\hat{H}_{\text{eff}} = \sum_{a=1}^N \hat{H}_a^\circ + \frac{1}{2}\sum_{a=1}^N\sum_{b\neq a}^N \hat{H}_{ab} \quad (6)$$

where $\hat{H}_a^\circ$ is the Hamiltonian of molecule a in the gas phase (eq 7) and $\hat{H}_{ab}$ is the QM/MM interaction Hamiltonian between molecules a and b with a treated quantum mechanically and b represented by partial atomic charges (or in general, multipoles) (eq 8).

$$\hat{H}_a^\circ = -\frac{1}{2}\sum_{i=1}^{2M}\nabla_i^2 - \sum_{\alpha=1}^A\sum_{i=1}^{2M}\frac{Z_\alpha(a)}{R_{\alpha i}(a,a)} + \sum_{i=1}^{2M}\sum_{j>i}^{2M}\frac{1}{r_{ij}} + E_{\text{nuc}} \quad (7)$$

$$\hat{H}_{ab} = -\sum_{b\neq a}^N\sum_{\beta=1}^A\sum_{i=1}^{2M}\frac{q_\beta(b)}{R_{\beta i}(b,a)} + \sum_{b\neq a}^N\sum_{\beta=1}^A\sum_{\alpha=1}^A\frac{Z_\alpha(a)\,q_\beta(b)}{R_{\alpha\beta}(a,b)} + E_{ab}^{\text{vdW}} \quad (8)$$

In eqs 7 and 8, r_{ij} are distances between electrons in a molecule, $R_{\alpha i}(a,b)$ are the distances between the nuclei of molecule a and the electrons in molecule b, $R_{\alpha\beta}(a,b)$ represent the distances between two nuclei in molecules a and b, $Z_\alpha(a)$ are nucleus charges of molecule a, and $q_\beta(b)$ are partial atomic charge on

Empirical Potential for Liquid Simulations

J. Phys. Chem. B, Vol. 101, No. 4, 1997 **659**

atoms in molecule b. The van der Waals term in eq 8 was necessary to describe short-range electron repulsions and dispersion interactions between the QM and MM regions.[33,43] The van der Waals term is represented in terms of the Lennard-Jones potential

$$E_{ab}^{\text{vdW}} = \sum_{\alpha=1}^{A} \sum_{\beta=1}^{A} 4\epsilon_{\alpha\beta} \left[\left(\frac{\sigma_{\alpha\beta}}{R_{\alpha\beta}(a,b)} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{R_{\alpha\beta}(a,b)} \right)^{6} \right] \quad (9)$$

where $\sigma_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$ are empirical parameters which can be derived using standard combining rules such that $\sigma_{\alpha\beta} = (\sigma_\alpha\sigma_\beta)^{1/2}$ and $\epsilon_{\alpha\beta} = (\epsilon_\alpha\epsilon_\beta)^{1/2}$. σ_α and ϵ_α are parameters to be determined later, which depend on the atomic number and sometimes also on its hybridization.[25,44]

With the above expressions, eq 6 may be rewritten as follows:

$$\hat{H}_{\text{eff}} = \sum_{a=1}^{N} \hat{H}_a \quad (10)$$

with

$$\hat{H}_a = \hat{H}_a^{\circ} + \frac{1}{2}\sum_{b \neq a}^{N} \hat{H}_{ab} \quad (11)$$

Note that eq 11 contains electronic coordinates of molecule a only and is similar to a solution system in which the solute is treated by a QM method and the solvent molecules are described by MM force fields. It includes interactions with all other solvent molecules in the system and is used to determine the wave function in variational QM calculations for each solvent molecule.

Within the Born–Oppenheimer approximation, the total energy of the system, E_{tot}, is the expectation value of the wave function Φ over the electronic Hamiltonian, and it is a function of the nuclear positions $\{\mathbf{R}_\alpha(a)\}$, and the MO coefficients $\{c_{\mu i}(a)\}$:

$$E_{\text{tot}} = \langle \Phi | \hat{H}_{\text{eff}} | \Phi \rangle - N \langle \Phi^{\circ} | \hat{H}^{\circ} | \Phi^{\circ} \rangle$$
$$= \sum_{a=1}^{N} E_a + \frac{1}{2}\sum_{a=1}^{N}\sum_{b \neq a}^{N} \epsilon_{ab} - NE^{\circ} \quad (12)$$

where E_a is the energy of molecule a in solution, E° is the energy of an isolated molecule in the gas phase, and ϵ_{ab} is the interaction energy between molecules a and b. In contrast to molecular mechanics potentials, the energy of a single gas phase molecule is nonzero in QM calculations. Thus, this constant gas phase energy is subtracted from the energy expression in eq 12 so that the energy zero corresponds to a gas of molecules at infinite separation.

For the liquid state, E_a is determined by eq 13:

$$E_a = \langle \Psi_a | \hat{H}_a^{\circ} | \Psi_a \rangle \quad (13)$$

We note that Ψ_a is the molecular wave function for molecule a in solution, which minimizes the energy of the Hamiltonian of eq 11. It must be distinguished from the wave function of the molecules in the gas phase, Ψ°, which minimizes the energy of the Hamiltonian of $\hat{H}^{\circ}$ alone (note that all solvent molecules have identical gas phase Hamiltonians):[25]

$$E^{\circ} = \langle \Psi^{\circ} | \hat{H}^{\circ} | \Psi^{\circ} \rangle \quad (14)$$

According to the variational principle, E_a is always greater than or equal to E_a°. The interaction energy between molecules a and b, ϵ_{ab}, is defined by eq 15:

$$\epsilon_{ab} = \frac{1}{2}[\langle \Psi_a | \hat{H}_{ab} | \Psi_a \rangle + \langle \Psi_b | \hat{H}_{ba} | \Psi_b \rangle] \quad (15)$$

Thus, it is clear that $\epsilon_{ab} = \epsilon_{ba}$, although the two individual terms in eq 15 do not always give identical values due to the imbalance introduced by the partitioning of a molecular system into a QM region and an MM region.[33] The definition of eq 15 eliminates such an imbalance; i.e., the interaction energy for a dimer complex is not always identical when one molecule is treated quantum mechanically and the other classically, and vice versa. We note that analogous expressions have been used in quantum chemical calculations including the extended Hückel theory designed for geometry optimizations.[45]

Here, the computational procedure should be mentioned. For each molecule, its wave function is determined by standard HF-SCF procedures, with the inclusion of the environmental effects through the QM/MM interaction Hamiltonian defined by eq 11. The computational details involve a modification of the Fock matrix elements by including additional one-electron integral terms, and they have been described in previous publications.[31–35] The resulting wave function is then used to derive the multipole expansion or, simply, partial atomic charges as it is preferred here. Obviously, the way in which these point charges are determined brings about a crucial approximation in the present method. Further discussion will be presented in the next section. However, a key point in deriving the MODEL potential is that the charge projection procedure must be optimized or parametrized by reproducing experimental liquid properties through Monte Carlo or molecular dynamics simulations. Because the optimization of the charge projection procedure is purely aimed at reproducing experimental thermodynamic and structural data of the liquid, the MODEL potential must be regarded as an empirical model. However, errors due to the use of semiempirical methods to describe intermolecular interactions, and a lack of explicit treatment of electron correlation effects as well as the use of a hybrid QM/MM technique, may be alleviated or eliminated through empirical parametrization. Like molecular mechanics, the MODEL potential uses experimentally derived parameters to strive for accuracy. This is indeed the spirit taken by all semiempirical QM models, though it has only been applied to isolated (gas phase) systems.[38–40]

The forces on nuclear positions are determined by differentiating eq 12 with respect to nuclear coordinates:

$$\mathbf{F}_\alpha(a) = -\nabla_\alpha(a)E_{\text{tot}} \quad (16)$$

Expressions for the coordinate derivatives may be found in early studies.[27,28,46] Consequently, geometry minimization and standard molecular dynamics simulations may be carried out using the MODEL potential. Alternatively, Car–Parrinello type molecular dynamics calculations,[26] which has been extended to Hartree–Fock MO simulations with the use of the NDDO approximation by Field,[28] can be performed using the MODEL potential for fluid systems. Although a detailed discussion is beyond the scope of this paper and the method will be described in a second paper, we only point out here that a Lagrangian for the system can be defined, subject to the constraint terms of eq 5, to generate the equations of motion for molecular dynamics simulations:

$$L = \frac{1}{2}m_c\sum_{a=1}^{N}\sum_{\mu,i}\left[\frac{dc_{\mu i}(a)}{dt}\right]^2 + \frac{1}{2}\sum_{a=1}^{N}\sum_{\alpha=1}^{A}m_\alpha\left[\frac{dR_\alpha(a)}{dt}\right]^2 -$$
$$E_{\text{tot}}[\{R_\alpha(a)\},\{c_{\mu i}(a)\}] \quad (17)$$

where m_c is the "mass" of the MO coefficients, m_α is the mass of atom α, and t is the time.

3. Computational Details

Clearly, the key to success of the MODEL potential is to find an appropriate procedure for determining the point charges $\{q_\alpha(a)\}$ from the molecular wave functions $\{\Psi_a\}$ as well as the optimization of Lennard-Jones parameters in eq 9. Conventional HF-SCF methods can be used to obtain $\{\Psi_a\}$, except that, after each iteration, a new set of charges will be calculated and used in constructing the next Fock matrix. Alternatively, electronic variables $\{c_{\mu i}(a)\}$, and thus $\{q_\alpha(a)\}$, may be determined dynamically using an algorithms analogous to that originally developed by Car and Parrinello and by Field.[26,28] However, the latter calculations must be performed at a much smaller time interval between neighboring points. There have been doubts as to whether the method would be more efficient than performing a full QM calculation with longer dynamic integration steps in trajectory calculations.

The simplest procedure for charge partitioning perhaps is the Mulliken population analysis,[47] in which the partial charges on atoms are defined, in the NDDO approximation, as follows:

$$Q_\alpha(a) = Z_\alpha - \sum_{i=1}^{M} \sum_{\mu}^{\text{on } \alpha} 2c_{\mu i}(a)\, c_{\mu i}(a) \qquad (18)$$

where M is the number of occupied molecular orbitals in molecule a. However, the quality of the Mulliken charges is known to be poor in describing electrostatic properties, especially in semiempirical QM calculations. Nevertheless, the semiempirical AM1 or MNDO Mulliken charges have been shown to have good correlation with partial charges obtained from ab initio Mulliken charges as well as electrostatic potential derived charges, suggesting that the relative charge separation is reasonably represented.[48,49] Consequently, it might be possible to simply scale the Mulliken population charges of neutral molecules for use in the MODEL potential:

$$q_\alpha(a) = K_m Q_\alpha(a) \qquad (19)$$

where K_m is a universal constant associated with a particular quantum chemical model, and its value should be optimized through comparison of the computed and experimental liquid properties. In the present paper, we decided to use the scaled Mulliken population charges to demonstrate the feasibility of the MODEL potential to describe interactions in bimolecular complexes. We found that a value of 2.2 for K_m was appropriate for the AM1 model to yield good agreement with ab initio MP2/6-31G(pd) results.

Another commonly used procedure for determining atomic charges is based on the fitting of classical molecular electrostatic potentials (MEPs) resulting from point charges to QM MEPs.[50] This procedure has been widely used in generating partial charges for molecular dynamics simulations.[51] However, semiempirical MEP-derived charges as in the case of Mulliken charges are generally too small in magnitude, but excellent correlation can be established with charges obtained from high-level ab initio wave functions.[48,49] Therefore, a similar scaling procedure may be proposed using semiempirical MEP-derived charges. The difference between the Mulliken charges and the MEP-derived charges is that the former retains the property of relative atomic electronegativities in the molecule, whereas the latter provides a better description of the electrostatic properties around the molecule.

Finally, another approach in generating "high"-level atomic partial charges with semiempirical wave functions is the charge model (CM) proposed by Cramer and Truhlar.[52] The charge model is a multilinear mapping procedure in which $\{q_\alpha^{\text{CM}}\}$ depend on all Mulliken population charges and on some of the bond orders. The CM model contains empirical parameters, thereby providing an opportunity to optimize these parameters to yield good results in comparison with experiments. These authors have shown that with the use of semiempirical AM1 or PM3 Hamiltonian the CM model can be parametrized to yield charges and molecular dipole moments in excellent agreement with experimental and high-level ab initio results.[52]

Below, we outline an algorithm for optimizing the molecular wave function $\{\Psi_a\}$ in full variational QM calculations.

a. Initialize the computational procedure by defining the initial values for the MO coefficients, $\{c_{\mu i}^{\circ}(a)\}$, and atomic coordinates, $\{\mathbf{R}_\alpha(a)\}$, of all solvent molecules. Determine the initial set of partial atomic charges, $\{q_\alpha^{\circ}(a)\}$, using these MO coefficients. During a Monte Carlo or molecular dynamics simulations, $\{c_{\mu i}^{\circ}(a)\}$ are typically the optimal orbital coefficients from the previous step.

b. Construct the Fock matrix for each individual solvent molecule, $\{F_{\mu\nu}(a)\}$, by adding the one-electron integral term resulting from the surrounding solvent atomic partial charges to the "gas phase" Fock matrix, $\{F_{\mu\nu}^{\circ}(a)\}$. This follows the exact scheme used in hybrid QM/MM methods.[31]

c. Diagonalize the N Fock matrices, each of which has a dimension of $N_b \times N_b$, to yield the energy, E_{tot}^m, and a new set of orbital coefficients, $\{c_{\mu i}^m(a)\}$. Here, the superscript m indicates the mth iteration in the self-consistent-field calculation.

d. Check for convergence of the density matrix or total energy of the system. If convergence has been achieved, proceed to the next minimization step or simulation step. Otherwise, calculate the new charge distribution, $\{q_\alpha^m(a)\}$, from $\{c_{\mu i}^m(a)\}$. Return to step b. The procedure proceeds as in conventional HF-SCF methods until self-consistency is achieved.

If the monomer geometry was held fixed during a calculation, the electron integrals only need to be computed once and stored in computer memory. This is possible due to the small size of each solvent molecule as opposed to the entire liquid system and the use of semiempirical methods in which numerous integrals are assumed to be zero.

Here, we note that the Fock matrix construction and diagonalization are performed for each individual solvent molecule, which involves a total of N "separate" electronic structure calculations at each iterative step. Thus, the quantum chemical procedure for the entire liquid system is strictly $O(N)$ scaling in computational effort and parallelizable. Of course, the normal HF-SCF scaling behavior for each individual solvent molecule still exists. This is in contrast to a HF calculation of the full liquid system, which would require the construction and diagonalization of the full Fock matrix in addition to the storage requirement of $O(N^2 \times N_b^2)$. For example, the dimension of the Fock matrix would be 1296 × 1296 for a box of 216 water molecules, in comparison with 216 smaller Fock matrices of 6 × 6 dimension in the MODEL potential. Several reports have appeared describing linear scaling methods for HF and DFT calculations of macromolecules, including novel methods for updating the density/orbitals without explicit diagonalization.[53−58] It is not yet clear whether these approaches are viable for use in fluid simulations. The method proposed here, which employs a hybrid QM/MM technique and involves empirical parametrization, appears to be a practical, alternative approach for molecular dynamics and Monte Carlo calculations.

The other empirical parameters in the MODEL potential,

Empirical Potential for Liquid Simulations

J. Phys. Chem. B, Vol. 101, No. 4, 1997 **661**

TABLE 1: Parameters Used in the MODEL Potential

atom	σ (Å)	ϵ (kcal/mol)	atom	σ (Å)	ϵ (kcal/mol)
C	3.40	0.08	N	3.25	0.17
H (on C)	2.00	0.01	O(sp^2)	3.50	0.20
H (on heteroatoms)	0.80	0.05	O(sp^3)	3.10	0.15

TABLE 2: Computed Energetic and Structural Results for Bimolecular Complexes
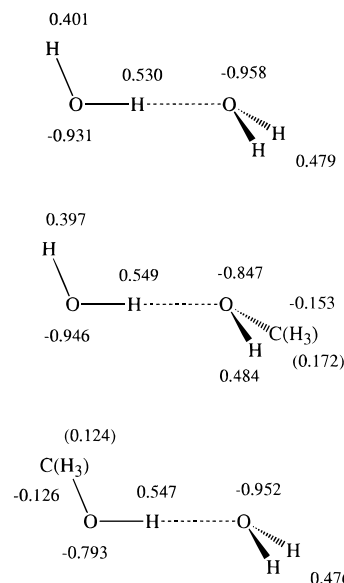
	MP2/6-31G(pd)			MODEL		
complex	E (kcal/mol)	R (Å)	θ (deg)	E (kcal/mol)	R (Å)	θ (deg)
HOH···OH$_2$	−5.47	1.97	157	−5.28	1.86	156
HOH···OHCH$_3$	−5.27	1.94	118	−5.84	1.83	150
CH$_3$OH···OH$_2$	−5.47	1.94	116	−4.91	1.86	163
HOH···O=CH$_2$	−4.81	2.03	108	−4.80	2.04	136
HCONH$_2$···OH$_2$ (cis)	−7.80	1.99	103	−6.85	1.84	132
HCONH$_2$···OH$_2$ (trans)	−6.42	2.03	140	−7.76	1.83	179

Figure 1. Mulliken populations charges after scaling by K_m in bimolecular complexes.

aside from the semiempirical quantum chemical model itself, include the Lennard-Jones parameters, σ and ϵ. These parameters may also be determined through a series of fluid simulations by fitting the computed structural and thermodynamic properties of the liquid to experiment. This approach was used by Jorgensen and co-workers in developing their optimized potentials for liquid simulations (OPLS) and has been adopted by other groups to generate empirical potential functions.[59,60] Based on prior experience, a single set of Lennard-Jones parameters for each atom or each hybridization type might be sufficient; however, careful optimization of these parameters would be necessary for use with the MODEL potential.

4. Results and Discussion

To illustrate the method, sample calculations are presented on bimolecular complexes using the AM1 semiempirical Hamiltonian.[39,61] The results are obtained by optimizing the universal scaling constant K_m, and the van der Waals parameters by fitting the interaction energies and hydrogen-bonding distances predicted using the MODEL potential to those from ab initio MP2/6-31G(dp) results. The purpose of this calculation is to demonstrate that the MODEL potential can be parametrized to reproduce a given set of target properties, and consequently, it may be used for fluid simulations. The parameters used in the present calculation are summarized in Table 1. In bridging the semiempirical AM1 Hamiltonian with an MM force field, we have adopted the exact approach used in previous hybrid QM/MM potentials for solution systems. For additional details, the readers are directed to ref 31. In all calculations, monomer geometries are held fixed, and only the hydrogen-bond distance and angle are optimized for each complex. Experimental geometry was used for water, whereas the AM1 geometries were adopted for other molecules.

Computed energetic and structural results are listed in Table 2. In these calculations, the quantum chemical model (the AM1 Hamiltonian) was fixed without further adjustments, whereas the Lennard-Jones parameters and the charges scaling constant were varied to obtain a good fit to the ab initio data. Overall, the results predicted with the MODEL potential are in reasonable agreement with the target values, particularly for simple cases such as the water dimer. However, for the methanol/water complex the results in Table 2 revealed fundamental problems associated with the semiempirical model. It was not possible to obtain good agreement with ab initio calculations for both hydrogen-bond donor and acceptor complexes of methanol with water. This, in fact, is an intrinsic problem of the AM1 model (as well as the MNDO and PM3 models). Full AM1 calcula-

tions yield an association energy of −2.7 kcal/mol for the donor complex (CH$_3$OH···OH$_2$), whereas the interaction energy is −4.8 kcal/mol for the acceptor complex (HOH···OHCH$_3$).[39] On the contrary, MP2/6-31G(dp) calculations give interaction energies of −5.5, and −5.3 kcal/mol, respectively. Clearly, the AM1 Hamiltonian severely underestimates hydrogen-bonding interactions, and the order of interaction energies of donor and acceptor complexes is reversed. The difficulty is reflected by the results predicted with the MODEL potential in Table 2. Thus, it would be desirable to further optimize the semiempirical QM parameters in deriving the MODEL potential, with special consideration of intermolecular interactions to appropriately account for the relative electronegativity of different atoms in organic systems. Such problems do not appear to be present in hybrid QM/MM calculations using ab initio QM methods even at a modest HF/3-21G level as demonstrated in a recent study by our group.

To illustrate the mutual charge polarization of monomers in a bimolecular complex, computed Mulliken population charges, after scaling by K_m, are shown in Figure 1 for selected complexes. In each case, polarization effects are significant as demonstrated by an increase in the magnitude of partial charges in comparison with those of the isolated molecules. The anisotropy of charge distribution in asymmetric orientations was also clearly reflected.

There are several issues that should be discussed regarding the present MODEL potential. In essence, the MODEL potential is very similar to Berne's fluctuating charge model.[15] However, the MODEL potential is based on quantum chemical methods with "fluctuating charges" projected from the molecular wave functions. On the other hand, Berne's FC model is entirely an empirical model. Furthermore, in the FC model, the Coulomb interaction for intramolecular pairs is taken to be the Coulomb overlap integral between Slater orbitals,[15] which was introduced by Rappé and Goddard.[23] The one-center limit was determined by varying the value of the orbital exponent. In the MODEL potential, intramolecular pair interactions are completely determined by the semiempirical quantum chemical method.[39] For intermolecular atomic pair interactions, the Coulomb law was directly applied in the FC model, whereas the MODEL potential adopted the approach introduced in the MNDO method for treating core−core repulsion terms.[39] Thus, the core−core interactions are made proportional to the two-

electron integrals of s-type orbitals centered on the nuclei. In such a two-electron term, the one-center limit was approximated by the Kolpman−Ohno type scheme in the Dewar−Thiel algorithm.[41] Consequently, the electron correlation effects are also empirically included, to a certain extent, in the hybrid QM/MM approach.

In the FC model, the charge equalization scheme can be applied, in principle, to the entire fluid system. In practice, the charge transfer between distant molecules will be severely overestimated since no barrier is present in the model to prevent charge equilibration through tunneling. Therefore, charge equilibration is restricted to intramolecular charge transfer in the FC model.[15] The MODEL potential as it is described in this paper is also restricted to only allow intramolecular charge transfer. However, intermolecular charge transfer may be introduced into the MODEL potential without the problems encountered in the FC model. This would be an important extension of the MODEL potential to protein systems and will be addressed in a later publication.

Several reports have appeared, describing "linear scaling" methods to the treatment of macromolecules including proteins.[53−58] For practical, dynamics simulation purposes, the most promising approaches appear to be those based on semiempirical models and approximate methods in constructing the total density matrix, e.g., the orbital localization scheme and Yang's divide-and-conquer algorithm.[56−58] The present MODEL potential differs from these approaches in that no attempt is made to construct the full density matrix for the entire system. This is because even when the density matrix can be precisely obtained, semiempirical models are not sufficiently accurate to describe intermolecular interactions necessary for fluid simulations.[38−40]

Can hybrid QM/MM methods embedded in the MODEL potential provide an accurate description of intermolecular interactions of polar and nonpolar systems in solutions and proteins? Indeed, we have found that hybrid QM/MM approaches making use of the AM1 Hamiltonian can yield good results for bimolecular complexes and free energy of solvation of organic molecules.[31,42] The success stems from a careful optimization of the empirical parameters that are inherently involved in hybrid QM/MM methods.[25,44] Consequently, electron correlation effects can be addressed and included in an average way in the parameterization procedure by fitting computed results to experimental data. This approach has been the basis for the parametrization of semiempirical models as well as for the development of empirical potential functions.[38−40,59,60]

5. Summary

This paper has presented a method for describing many-body polarization effects in fluid systems, making use of hybrid QM/MM techniques and semiempirical NDDO HF wave functions. In this approach, the electronic structure of each solvent molecule is represented by an antisymmetric determinant wave function, which can be determined by fully converged variational QM calculations at each step of the classical trajectory or can be treated as dynamic variables along with the nuclear coordinates in molecular dynamics simulations. In determining the molecular wave functions for the solvent molecules, hybrid QM/MM potentials are employed, in which empirical parameters, including the universal charge scaling constant and the Lennard-Jones parameters, are subjected to optimization to reproduce experimental thermodynamic properties for the liquid. Consequently, the electron correlation effects that are not explicitly included in the QM calculations are taken into account through

the parametrization procedure. The present MODEL potential is analogous in spirit to the FC model introduced by Berne and co-workers; however, the MODEL potential is a quantum chemical model as opposed to a purely empirical charge equilibration scheme employed in the FC model.

The capability of the MODEL potential for describing bimolecular interactions was illustrated by comparison of the results predicted using the MODEL potential with those obtained from full ab initio MP2/6-31G(dp) calculations. Reasonable agreement in energy and structure for bimolecular complexes was obtained, especially for simple dimer systems. This indicates that the MODEL potential may be optimized to reproduce liquid properties. The results of these studies will be reported in a forthcoming paper.

References and Notes

(1) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: London, 1987.

(2) (a) McCammon, J. A.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987. (b) Brooks, C. L., III.; Karplus, M.; Pettitt, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1.

(3) Burkert, U.; Allinger, N. L. *Molecular Mechancis*; ACS Monographs, Vol. 177, American Chemical Society: Washington, DC, 1982.

(4) Stillinger, F. H.; David, C. W. *J. Chem. Phys.* **1978**, *69*, 1473.

(5) Barnes, P.; Finney, J. L.; Nicholas, J. D.; Quinn, J. E. *Nature* **1979**, *282*, 459.

(6) van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. J. *J. Mol. Biol.* **1987**, *198*, 721.

(7) Rullman, J. A. C.; van Duijnen, P. Th. *Mol. Phys.* **1988**, *63*, 451.

(8) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556.

(9) Ahlstrom, P.; Wallqvist, A.; Engstrom, S.; Jonsson, B. *Mol. Phys.* **1989**, *68*, 563.

(10) Caldwell, J. W.; Dang, L.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 9144. (b) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208.

(11) Niesar, U.; Corongiu, G.; Clementi, E.; Kneller, G. R.; Bhattacharya, D. K. *J. Phys. Chem.* **1990**, *94*, 7949.

(12) Warshel, A. *J. Phys. Chem.* **1990**, *94*, 460.

(13) Straatsma, T. P.; McCammon, J. A. *Chem. Phys. Lett.* **1991**, *177*, 433.

(14) (a) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180. (b) Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1995**, *99*, 6374. (c) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Comput. Chem.* **1995**, *16*, 1141.

(15) (a) Rick, S. E.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141. (b) Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. J. *J. Mol. Liq.* **1995**, *65*, 31. (c) New, M. H.; Berne, B. J. *J. Am. Chem. Soc.* **1995**, *117*, 7172. (d) Rick, S. W.; Berne, B. *J. Am. Chem. Soc.* **1996**, *118*, 672. (e) Bader, J. S.; Berne, B. J. *J. Chem. Phys.* **1996**, *104*, 1293. (f) Stuart, S. J.; Berne, B. J. *J. Phys. Chem.* **1996**, *100*, 11934.

(16) (a) Gao, J.; Habibollahzadeh, D.; L. Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460. (b) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689.

(17) Böttcher, C. J. *Theory of Electric Polarization*; Elsevier: Amsterdam, 1973.

(18) (a) Applequist, J. B.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1972**, *94*, 2952. (b) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

(19) (a) Sanderson, R. T. *Science* **1951**, *114*, 670. (b) Sanderson, R. T. *Chemical Bonds and Bond Energy*; Academic Press: New York, 1976.

(20) (a) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. *J. Chem. Phys.* **1978**, *68*, 3801. (b) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford Univeristy Press: Oxford, 1989.

(21) (a) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219. (b) Mortier, W. J.; Genechten, K. V.; Gasteiger, J. *J. Am. Chem. Soc.* **1985**, *107*, 829.

(22) No, K. T.; Grant, J. A.; Scheraga, H. A. *J. Phys. Chem.* **1990**, *94*, 4732.

(23) Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.

(24) York, D. M. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1995**, *29*, 385. (b) York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 159.

(25) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

(26) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2473.

(27) Head-Gordon, M.; Pople, J. A. *J. Phys. Chem.* **1988**, *92*, 3063.

(28) (a) Field, M. J. *Chem. Phys. Lett.* **1990**, *172*, 83. (b) Field, M. J. *J. Phys. Chem.* **1991**, *95*, 5104. (c) Field, M. J. *J. Chem. Phys.* **1992**, *96*, 4583.

(29) (a) Tuckerman, M. E.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103*, 150. (b) Carrloni, P.; Blochl, P.; Parrinello, M. *J. Phys. Chem.* **1995**, *99*, 1338. (c) Laasonen, K.; Klein, M. L. *J. Am. Chem. Soc.* **1994**, *116*, 11620. (d) Shah, R.; Payne, M. C.; Lee, M.-H.; Gale, J. D. *Science* **1996**, *271*, 1395. (e) Nusterer, E.; Blochl, P. E.; Schwarz, K. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 175.

(30) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213.

(31) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 7, pp 119−185.

(32) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(33) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(34) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

(35) (a) Topia, O.; Colonna, F.; Angyan, J. G. *J. Chim. Phys. Phys.-Chim. Biol.* **1990**, *87*, 875. (b) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374. (c) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Phys. Chem.* **1993**, *97*, 11868. (d) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580. (e) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386.

(36) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; McGraw-Hill: New York, 1989.

(37) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(38) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(39) Dewar, M. J. S.; Zeobisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(40) Stewart, J. J. P. *J. Comput. Chem.***1989**, *10*, 209, 221.

(41) Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta* **1977**, *46*, 89.

(42) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.

(43) Floris, F.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784.

(44) Gao, J. In *Modeling the Hydrogen Bond*; Smith, D. A., Ed.; ACS Symposium Series 569; American Chemical Society: Washington, DC, 1994; pp 8−20.

(45) (a) Calzaferri, G.; Forss, L.; Kamber, I. *J. Phys. Chem.* **1989**, *93*, 5366. (b) Carbo, R.; Fornos, J. M.; Hernandez, J. A.; Sanz, F. *Int. J. Quantum Chem.* **1977**, *11*, 271. (c) Anderson, A.; B.; Hoffmann, R. *J. Chem. Phys.* **1974**, *60*, 4271.

(46) Dewar, M. J. S.; Yamaguchi, Y. *J. Comput. Chem.* **1978**, *2*, 25.

(47) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(48) (a) Merz, K. M., Jr. *J. Comput. Chem.* **1992**, *13*, 749. (b) Marrone, T. J.; Hartsough, D. S.; Merz, K. M., Jr. *J. Phys. Chem.* **1994**, *98*, 1341. (c) Ford, G. P.; Wang, B. *J. Comput. Chem.* **1993**, *14*, 1101. (e) Bakowies, D.; Thiel, W. *J. Comput. Chem.* **1996**, *17*, 87.

(49) (a) Gao, J.; Luque, F. J.; Orozco, M. *J. Chem. Phys.* **1993**, *98*, 2975. (b) Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1990**, *11*, 909.

(50) Momany, F. A. *J. Chem. Phys.* **1978**, *82*, 592.

(51) (a) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431. (b) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(52) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

(53) Li, X.-P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev.* **1993**, *B47*, 10891.

(54) Galli, M.; Parrinello, M. *Phys. Rev. Lett.* **1992**, *69*, 3547.

(55) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.

(56) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133.

(57) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1996**, *104*, 6643.

(58) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.,* in press.

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

(60) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(61) (a) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1. (b) Stewart, J. J. P. *QCPE 455*, **1983**, *455*.