JCTC Journal of Chemical Theory and Computation

Design of a Next Generation Force Field: The X-POL Potential

Wangshen Xie[†] and Jiali Gao*[,†,‡]

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455, and Centro Nacional de Supercomputación, Programa Biología Computacional, C/ Jordi Girona 29, 08034 Barcelona, Spain

Abstract: An electronic structure-based polarization method, called the X-POL potential, has been described for the purpose of constructing an empirical force field for modeling polypeptides. The X-POL potential is a quantum mechanical model, in which the internal, bonded interactions are fully represented by an electronic structure theory augmented with some empirical torsional terms. Nonbonded interactions are modeled by an iterative, combined quantum mechanical and molecular mechanical method, in which the molecular mechanical partial charges are derived from the molecular wave functions of the individual fragments. In this paper, the feasibility of such and electronic structure based force field is illustrated by small model compounds. A method has been developed for separating a polypeptide chain into peptide units, and its parametrization procedure in the X-POL potential is documented and tested on glycine dipeptide. We envision that the next generation of force fields for biomolecular polymer simulations will be developed based on electronic structure theory, which can adequately define and treat many-body polarization and charge delocalization effects.

1. Introduction

Molecular mechanics or force fields, employing empirical potential energy functions,[1] play a central role in computer simulations, which ultimately determine the accuracy of computational results.[2] Although remarkable success has been achieved in molecular dynamics and Monte Carlo simulations of liquids, solutions, and biopolymers, thanks to the development and careful validation of several force fields,[3–8] there are also a number of deficiencies in the current generation of force fields. First, the choice of the energy terms and the associated degrees of freedom are arbitrary in defining a force field.[9] As a result, different force fields often have seemingly very different parameters (for example, partial charges), but the computed dynamic and thermodynamic properties can be similar. Second, by and large, most force fields make use of the harmonic approximation to bond stretch and angle bending, and there is a lack of consideration of the coupling among different energy terms. Of course, a number of force fields do include anharmonicity and cross terms,[7,8,10,11] but there is no systematic way of improving the functional form and its performance because little is known about their effects in biomolecular simulations. Empirical force fields can be developed by parametrizing against observed vibrational frequencies.[1,9,11] Third, the treatment of many-body polarization effects suffers from difficulties in choosing the functional form and empirical parameters.[8,12–21,29] Just as atomic partial charges, there is no rigorous definition of atomic (or group) polarizabilities, nor is it measurable experimentally. There are many ways of describing molecular polarization, which of course is a well-defined quantity. Fourth, charge-transfer effects are ignored in all empirical force fields, and there is no obvious way of including these effects.[22–25] Although the energy contributions are typically small in most applications, charge transfer can be important in certain situations.[23,24] Finally, the form of the empirical potentials is not appropriate for modeling chemical reactions involving bond formation and bond breaking or regions significantly away from the adiabatic ground state. Although specialized potentials[26a] or general functional forms[26b] can be developed to treat specific

---

* Corresponding author e-mail: gao@chem.umn.edu.
† University of Minnesota.
‡ Centro Nacional de Supercomputación.

Design of a Next Generation Force Field

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **1891**

processes, they are restricted to applications to that particular case only. In view of these difficulties, it is of interest to consider alternative approaches to design a force field for biomolecular simulation and modeling that takes these effects into account. It is our hope that the present paper may contribute to this goal.

Combined quantum mechanical and molecular mechanical (QM/MM) potentials,[27−30] in principle, provide a reasonable solution to all the deficiencies mentioned above, at least for the region that is explicitly treated by quantum mechanics. The internal energy terms used in the force field as well as electrostatic interactions are described by the quantum chemical method used to represent the "QM" region. There is no ambiguity in the functional form, nor in the selection of internal degrees of freedom. Molecular polarization and charge transfer are naturally represented by electronic structure theory. And, of course, such a method can be used to model chemical reactions.[31] Ten years ago, we described a method for treating many-body polarization effects in fluid simulations by combined QM/MM techniques.[32] This method takes a different approach to treat polarization and electronic effects which do not have the problems in the classical polarizable force fields noted above.[8,12−21,29] We had envisioned developing an electronic structure based polarization force field, which hereafter is called the X-POL potential, for biomolecular simulations that naturally includes molecular polarization. In paper 1, we outlined the general principles and formalisms of this approach,[32] and in the second paper, we demonstrated the feasibility of carrying out fluid simulations with such a polarizable potential function for liquid water.[33] This approach was also applied to liquid HF.[34] In this paper, we develop the theory for constructing a force field to treat polypeptides on the basis of the X-POL potential.

We emphasize that the goal here is the development of a molecular-orbital based force field, which is empirical and contains parameters; these parameters shall be optimized to reproduce the properties of a set of target molecular systems following the same philosophy in developing and validating molecular mechanics force fields. Yet, the fundamental elements including the bonded and electrostatic "terms" are determined by electronic structure theory. We do not aim at developing a linear scaling method for a fully solvated protein system,[35−37] although such a treatment is esthetically appealing and the X-POL potential can be developed into a quantum model for biopolymers. This indeed is the approach undertaken in ab initio molecular dynamics and Car−Parrinello molecular dynamics which have been successfully used in numerous applications.[38−40] However, these studies are limited to small systems with relatively short simulations. It would be difficult in the near future to extend the method to much larger systems such as a solvated protein or nucleic acid. Semiempirical models overcome the problem of computational costs,[41−44] but they are considered inaccurate for general applications. Some time ago, Head-Gordon estimated that a Hartree−Fock (HF) or density functional theory (DFT) calculation of a protein of about 10 000 atoms would only be remotely feasible with the assumption of 100-fold increase in computer speed and a true linear scaling in molecular

size.[37] Such calculations in a single energy evaluation are interesting, which still represent a daunting task today. However, it is necessary to perform statistical mechanical simulations to obtain ensemble averages in order to compare the computed results with experiments. The need to repeat energy and gradients evaluations for millions of times is the computation bottleneck in electronic structure methods for condensed phase systems.

Clearly, the only possibility that HF or DFT methods can be used as a reliable force field in biomolecular simulations is to introduce approximations, keeping in mind that it is necessary to be able to evaluate the energy and gradients for a given configuration within a few seconds of time. To this end, an X-POL potential, which was also called the molecular-orbital derived empirical potential for liquids (MODEL),[32] has been constructed for simulations of liquid water and hydrogen fluoride,[33,34] in which three approximations were made: (a) the wave function of the entire system is constructed as a Hartree product of the antisymmetric wave functions of individual molecules, (b) the interactions between any pair of residues are evaluated by combined QM/MM techniques, and (c) the electronic structure of individual molecules are treated by a semiempirical HF model. In this method, the electronic structure of each solvent molecule or amino acid residue in a polypeptide is influenced and polarized by the electrostatic field generated from the rest of the system, which in turn affects the wave functions of other molecules. Consequently, the total energy of the system is determined self-consistently. Clearly, many-body polarization effects are naturally described by electronic structure theory. The main advantage of this approach is that the treatment of molecular polarization as well as other energy components can be systematically improved by using ab initio HF, DFT, or advanced techniques such as perturbation, multiconfiguration, and couple cluster theories.

The X-POL potential also differs from combined QM/MM approaches that treat the induced dipoles classically in the MM region.[28,29,45−50] The difficulties and uncertainty of representing polarization in classical force field[8,12−21,29] still exists in these coupled QM/MM-pol methods, whereas the X-POL potential treats the entire system equally. Furthermore, charge-transfer effects can easily be included in the X-POL potential, which would be exceedingly difficult in the classical treatment. The X-POL force field is designed as a quantum mechanical model for biomolecular simulations.

In what follows, we first review the X-POL potential for treatment of liquid systems without covalent-bond connection between molecules. Then, in section 3, we introduce the new theory for treating polypeptides in which covalent bonding connections between residues must be separated. Section 4 presents the algorithm and computational details, and section 5 highlights the results and parametrization process. Finally, we summarize the major findings of this study and future perspectives.

2. Theoretical Background

For completeness, this section reviews the method presented in refs 32 and 33. The next section contains the new

methodology developed in the present work. We first consider a system consisting of N molecules that are not covalently connected in a primary unit cell with periodic boundary conditions along with nearest image convention. For the sake of brevity, we assume this is a simple liquid system with identical solvent molecules such as liquid water; obviously there is no restriction for solutions and mixed solvents in the method presented below.[33] To focus our discussion, we assume that the readers are familiar with combined QM/MM methods. We make the first assumption that the wave function of the liquid system (Φ) is represented by a Hartree product of the antisymmetric wave functions of the individual molecules, $\{\Psi_I; I = 1, \cdots, N\}$

$$\Phi = \prod_{I=1}^{N} \Psi_I \tag{1}$$

where the individual molecular wave function is written as a Slater determinant of M doubly occupied molecular orbitals (MOs), $\{\phi_i(I)\}$ with $2M$ electrons in each molecule. As usual, the MOs are linear combinations of an atomic orbital basis set, $\{\chi_\mu\}$, spanning over the entire molecule, which are subjected to the orthonormal constraint

$$\Lambda_{ij}(I) = \sum_{\mu\nu} c_{i\mu}(I) c_{j\nu}(I) S_{\mu\nu}(I) - \delta_{ij} = 0 \tag{2}$$

where $S_{\mu\nu}(I)$ is the overlap integral between atomic orbitals χ_μ and χ_ν in molecule I.

The assumption made in eq 1 neglects the exchange correlation interactions between molecules, thus, the entire system does not satisfy the Pauli exclusion principle, but this approximation is quite reasonable in the spirit of a force field development. To account for the short-range exchange repulsion and the long-range dispersion interactions, we use an empirical function to parametrically model these effects, and we adopt the popular Lennard-Jones potential

$$E_{IJ}^{\text{vdW}} = \sum_{\alpha=1}^{A}\sum_{\beta=1}^{B} 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{R_{\alpha\beta}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{R_{\alpha\beta}}\right)^{6}\right] \tag{3}$$

where A and B are the number of atoms in molecules I and J, which are the same in this discussion, and the parameters $\epsilon_{\alpha\beta}$ and $\sigma_{\alpha\beta}$ can be derived using the combining rules such that $\epsilon_{\alpha\beta} = (\epsilon_\alpha\epsilon_\beta)^{1/2}$ and $\sigma_{\alpha\beta} = (\sigma_\alpha + \sigma_\beta)/2$. ϵ and σ are atomic empirical parameters that are considered to have the same meaning and treatment as in a typical MM force field, and they depend on the specific functional type.

The Hamitonian of the system can be written as

$$\hat{H} = \sum_{I=1}^{N} \hat{H}_I^o + \frac{1}{2}\sum_{I=1}^{N}\sum_{J\neq I}^{N} \hat{H}_{IJ} \tag{4}$$

where $\hat{H}_I^o$ is the electronic Hamiltonian of an isolated molecule in the gas phase, and $\hat{H}_{IJ}$ describes the interactions between molecules I and J. The interaction Hamitonian can be expressed by eq 5:

$$\hat{H}_{IJ}(\Psi_J) = -\sum_{i=1}^{2M} V_i(\Psi_J) + \sum_{\alpha=1}^{A} Z_\alpha(I) V_\alpha(\Psi_J) + E_{IJ}^{\text{vdW}} \tag{5}$$

Here, $Z_\alpha(I)$ is the nucleus charge of atom α in molecule I, and $V_t(\Psi_J)$ is the electrostatic potential of molecular J at either the electronic ($t = i$) or nuclear ($t = \alpha$) positions of molecule I. The electrostatic potential due to molecule J is defined as follows

$$V_t(\Psi_J) = -\int \frac{d\mathbf{r}\rho_J(\mathbf{r})}{|\mathbf{r}_t - \mathbf{r}|} + \sum_{\beta=1}^{B} \frac{Z_\beta(J)}{|\mathbf{r}_t - \mathbf{R}_\beta(J)|} \tag{6}$$

where $\rho_J(\mathbf{r})$ is the electron density of molecule J, derived from the molecular wave function, $\rho_J(\mathbf{r}) = |\Psi_J(\mathbf{r})|^2$.

The total potential energy of the system is

$$E_{\text{tot}} = \langle\Phi|\hat{H}|\Phi\rangle - \sum_{I=1}^{N} E_I^o \tag{7}$$

where $E_I^o = \langle\Psi_I^o|\hat{H}_I^o|\Psi_I^o\rangle$ is the energy of molecule I in the gas phase with the wave function Ψ_I^o, which has a constant value and is used here purely for setting the zero of energy of the condensed phase system corresponding to that of infinitely separated or noninteracting species.

In principle, eq 7 can be determined by standard HF theory with or without optimization of the instantaneous molecular wave function Ψ_I in the presence of all other molecules. Of course, the method above is not restricted to HF theory and can be equivalently written in the form of DFT or any other electronic structure methods. For exampled, Wesolowski and co-workers have used a frozen density functional method for large systems without optimization of the electron density of individual fragments.[51,52] The fragmental molecular orbital method developed by Kitaura and co-workers allows for full optimization of the wave function.[53]

Without further approximation, it is necessary to compute the two-electron integrals arising from different molecules, which would be too expensive for a force field. Fortunately, this problem can be adequately treated by a combined QM/MM approach, which is the second assumption of the X-POL potential. Here, the electronic integral in eq 6 is expressed as a multipole expansion on molecule J. The two-electron two-center Coulomb integrals can also be evaluated in exactly the same way as that described by Dewar and Thiel in semiempirical NDDO methods.[54] Alternatively, if we only use the monopole term, i.e., partial atomic charges, the interaction Hamiltonian can be simplified to

$$\hat{H}_{IJ}(\Psi_J) = -\sum_{i=1}^{2M}\sum_{\beta=1}^{B} \frac{e \cdot q_\beta(\Psi_J)}{|\mathbf{r}_i - \mathbf{R}_\beta(J)|} + $$
$$\sum_{\alpha=1}^{A}\sum_{\beta=1}^{B} \frac{Z_\alpha(I)q_\beta(\Psi_J)}{R_{\alpha\beta}} + E_{IJ}^{\text{vdW}} \tag{8}$$

where $q_\beta(\Psi_J)$ is the partial atomic charge on atom β of molecule J, fitted to reproduce the electrostatic potential of eq 6 from the wave function Ψ_J, and $R_{\alpha\beta}$ is the distance between two atoms. Previously, we have shown that intermolecular interactions can be adequately described simply by scaling the Mulliken population charges in the simulation of liquid water.[33] In developing an X-POL force field for

Design of a Next Generation Force Field

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **1893**

biopolymers, it is desirable to include at least the dipolar expansion terms.

With this treatment, the potential energy of eq 7 is consistently optimized to obtain the ground-state potential energy of the system.

Recently, Gascon et al.[55] described a self-consistent space-domain decomposition method for computing electrostatic potentials of proteins. The method reported in that work appears to be the same as that described above except that Morokuma's ONIOM and the 6-31G(d) basis set were used.[56,57] Surprisingly, these authors do not appear to be aware of the method reported in refs 32−34. Soon after the publication of ref 32, Field also described a similar implementation making use of both the AM1 and HF/STO-3G method.[58]

3. The Electronic-Structure Polarization Force Field for Proteins

For biomolecular systems such as proteins and nucleic acids, the division of the entire system into individual molecular fragments is not obvious because each residue is covalently connected to its neighbors.[29,59−64] In this case, it is necessary to decide the basic unit for the "molecular partition" and to consider the effects of charge delocalization between neighboring fragments. In this section, we present a novel procedure for constructing a force field based on molecular orbital theory, for energy minimization and dynamics simulations of proteins.

3.1. Quantum Mechanical Model. Before we begin constructing the X-POL force field, a critical decision must be made on the choice of a specific quantum mechanical model to represent the system. Of course, it would be ideal to use an accurate electronic structure theory such as CCSD-(T) or a well-tested DFT model along with a large basis set. However, these methods are not practical in the foreseeable future, and it is not clear if DFT methods can yield accurate results without rigorous treatment of dispersion interactions. The most practical choice is semiempirical quantum mechanical models coupled with proper parametrizations, such as the self-consistent extended Huckel theory (SC-EHT),[66,67] and the formalisms based on the neglect diatomic differential overlap (NDDO) approximation.[41] The recent parametrization of the self-consistent charge density functional tight-binding (SCC-DFTB) model[68,69] yielded promising results; however, the procedure and the use of tabulated electronic integrals make it difficult for force field development. It would be of interest to improve the atomic parameters in the SC-EHT method. We find that the general formalisms used in the MNDO,[70] AM1,[42] and PM3[43] models are most appealing because the theory is well-defined and has been extensively tested. The semiempirical formalisms contain atomic parameters, and the total number of parameters are no more than those associated with a given atom type in the current empirical force fields.

We anticipate that the semiemipirical parameters will be fully optimized for each functional group in the amino acid residues, keeping in mind that we are interested in developing a *force field* rather than a "general" QM model. The parametrization will necessarily include optimizations of the
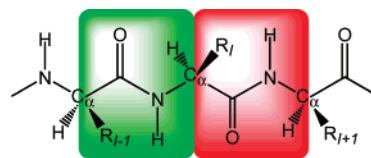

Figure 1. Definition of peptide units and the division of the C_α boundary atom. Two quantum mechanical fragments are highlighted in green and red, corresponding to residues *I-1* and *I*.

molecular geometries (including radial distribution functions), energies (such as heats of formation), electronic structural properties (such as molecular dipoles, electron affinities, and ionization potentials), spectroscopic data (for example, vibrational frequencies, NMR chemical shifts), and condensed phase properties (such as heats of vaporization, density, diffusion constants, relaxation time, and solvation free energies) among others for a set of selected compounds representing different functionalities. Of course, the X-POL force field can be systematically improved by increasing the level of the QM theory employed. It is expected that the parametrization process will become less dependent on fitting against experimental data (or high-level QM results) as the level of the QM model increases. In this paper, our focus is on the theory for constructing a polypeptide chain represented by a QM model. We adopt the Austin Model 1 (AM1)[42] method to demonstrate the method without further optimization for specific functional groups, except the boundary atoms discussed below.

3.2. System Partition and Boundary Definition. We consider a system of polypeptide of N residues, which is divided into N subunits or fragments. The interactions among different subunits are determined through a combined QM/MM algorithm. It would be natural to use the formal chemical structure of each amino acid residue as the "QM" subunit; however, it is more appropriate to keep intact the resonance delocalization in a peptide bond in electronic structure calculations. Thus, we adopt the "peptide unit" convention defined in the IUPAC nomenclature (*Pure Appl. Chem.* **1974**, *40*, 291−308. http://www.chem.qmul.ac.uk/iupac/misc/ppep1.html) which consists of the −CHR−CO−NH− atoms. For our computational purpose, which will become clear below, we make the sequence separation across the C_α atoms of adjacent residues, as illustrated in Figure 1. Thus, the Ith peptide unit contains the atoms $-C_\alpha{}^I R^I-CO-NH-C_\alpha{}^{I+1}H-$, in which $NH-C_\alpha{}^{I+1}H$ belongs to the $(I+1)$th amino acid residue. In our definition, the C_α atoms are equally *shared* by adjacent peptide units. In the following, we follow the IUPAC recommendation to simply refer the "peptide unit" as a "residue" when no ambiguity arises.

In this partition scheme, each residue (peptide unit) shares two C_α atoms with the neighboring residues, except the N- and C-termini. We call these atoms the boundary atoms (Figure 1). With the use of a semiempirical QM model, the boundary carbon atom has the standard valence s and p orbitals and four valence electrons. Adopting the generalized hybrid orbital (GHO) approach for the treatment of a QM-MM boundary in combined QM/MM calculations,[59,60] we make the same transformation of the sp atomic orbitals (AOs)
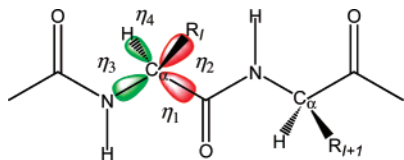
Figure 2. Assignment and sequence of hybrid orbitals on the boundary atom. Hybrid orbitals in red and green belong to different QM fragments.

on the boundary atom into a set of four orthonormal hybrid orbitals (HOs), on the basis of the local geometry about the C_α atom

$$\begin{pmatrix} \eta_1^I \\ \eta_2^I \\ \eta_3^I \\ \eta_4^I \end{pmatrix} = [\mathbf{T}_B(I)]^{-1} \begin{pmatrix} s_1^I \\ p_x^I \\ p_y^I \\ p_z^I \end{pmatrix} \qquad (9)$$

where the superscript I specifies that the orbitals are located on atom $C_\alpha(I)$, and the subscript B indicates that the transformation matrix has dimensions of (4×4) for the boundary atom. The transformation matrix in eq 9 is defined by the geometry of the four atoms bounded to the boundary atom and their local coordinates, and its expression has been given in ref 60. We note that the hybrid orbitals are orthonormal by construction, and they are also orthogonal to all other AO basis functions due to the NDDO approximation. In ab initio HF theory, the HOs need to be orthogonalized to the rest of the basis functions, and procedures have been described previously.[61,63]

In defining the HOs in eq 9, the transformation matrix, $\mathbf{T}_B(I)$, is constructed in such a way that the orientations of the four HOs are pointing sequentially toward the carbonyl carbon, the $C_\beta(I)$ (or $H_{\alpha2}$ for glycine) atom of the side chain, the amino nitrogen of the $(I\text{-}1)$th peptide unit, and the H_α atom (Figure 2). Therefore, there are two boundary atoms and four boundary hybrid orbitals in the Ith residue in the present QM partition: the first two hybrid orbitals, η_1^I and η_2^I, on the $C_\alpha(I)$ atom and the third and the fourth hybrid orbitals, η_3^{I+1} and η_4^{I+1}, on the $C_\alpha(I + 1)$ atom. Assuming that there are N_I atomic orbital basis functions on all other atoms in residue I, we have a total of $N_I + 4$ basis functions, called active orbitals, that are mixed atomic and hybrid orbitals to form the MOs of the subsystem. We note that, similar to the GHO method,[59–62] the remaining four hybrid orbitals, two from $C_\alpha(I)$ and two from $C_\alpha(I + 1)$, will also be used in constructing the Fock matrix, but they are not variationally optimized in self-consistent field (SCF) calculations of residue I.

3.3. Potential Energy Surface and Procedure. In optimizing the antisymmetric wave function Ψ_I, of residue I, we note that this QM subunit is embedded in the electric field of classical partial charges (or multipoles if higher order of density expansion is used) of the rest of the system, although they are also treated quantum-mechanically. Thus, in principle, the computational procedure is identical to that used in the GHO method for combined QM/MM systems,[59,60] except that two "active" hybrid orbitals from each boundary

atom participate in the SCF optimization. Specifically, if the density matrix for residue I is $\mathbf{P}^H(I)$, which has dimensions of $[N_I + 4] \times [N_I + 4]$, the total *interaction energy* between residue I and the rest of the system is

$$E_I = \sum_{\mu\nu} P_{\mu\nu}^H(I) H_{\mu\nu}^H(I) + 1/2 \sum_{\mu\nu} \sum_{\lambda\sigma} P_{\mu\nu}^H(I) P_{\mu\nu}^H(I) W^H(\mu\nu,\lambda\sigma) + E_{IP}^{nuc} + E_I^{nuc} - E_I^o \quad (10)$$

where the superscript H indicates that all matrix elements are given in the mixed AO and HO basis, $H_{\mu\nu}^H(I)$ is an element of the "effective" one-electron integral matrix that includes the interaction Hamiltonian of eq 8, $W^H (\mu\nu,\lambda\sigma)$ is the usual two-electron integrals including both Coulomb $(\mu\nu,\lambda\sigma)$ and exchange $(\mu\lambda,\nu\sigma)$ terms, and E_I^{nuc} and E_{IP}^{nuc} are the nuclear Coulombic energies within residue I and that with the rest of the protein system, respectively. The effective Hamiltonian matrix element $H_{\mu\nu}^H(I)$ is given below[59,60]

$$H_{\mu\nu}^H(I) = H_{\mu\nu}^{o,H}(I) + J_{\mu\nu}^H(I) + \frac{1}{2}\sum_{i,j=3}^{4} P_{\eta_i\eta_j}^H(I) W^H(\mu\nu,\eta_i^I \eta_j^I) + \frac{1}{2}\sum_{i,j=1}^{2} P_{\eta_i\eta_j}^H(I + 1) W^H(\mu\nu,\eta_i^{I+1} \eta_j^{I+1}) \quad (11)$$

where η_i specifies the boundary auxiliary orbitals from residues (I) and $(I +1)$ that are *not* optimized in the SCF for residue I, $H_{\mu\nu}^{o,H}(I)$ is the standard one-electron matrix for residue I, $J_{\mu\nu}^H(I)$ is the "QM/MM" one-electron integral due to the first term of eq 8 summed over all other residues other than I, and $P_{\eta_i\eta_j}^H(I)$ and $P_{\eta_i\eta_j}^H(I + 1)$ are the populations of the auxiliary hybrid orbitals specified by eq 11.

The total potential energy of the entire system is

$$E_{tot} = \frac{1}{2}\sum_{I=1}^{N} E_I \qquad (12)$$

We further define the interaction energy between residues I and J by[32,33]

$$E_{IJ} = \frac{1}{2}[<\Psi_I|\hat{H}_{IJ}|\Psi_I> + <\Psi_J|\hat{H}_{JI}|\Psi_J>] \qquad (13)$$

This is necessary to ensure that $E_{IJ} = E_{JI}$ because in combined QM/MM calculations the two integrals in the bracket parentheses may not be identical.

Although the wave function of eq 1 can be variationally optimized for all residues simultaneous in each SCF cycle, it is more convenient to sequentially optimize the wave function of each residue, by keeping the partial charges (derived from the corresponding wave functions) of the rest of the system fixed. Thus, we have a double iterative SCF procedure: (1) the SCF optimization of the wave function of each residue and (2) the SCF optimization of the mutual polarization of the entire system. Specifically, after the individual wave functions for all residues are converged, which constitute one iterative cycle in the "system" SCF, we check the convergence of the total energy of the system in eq 11. This is repeated until the total energy is converged to a given tolerance. This double-SCF procedure has been used previously in our treatment of liquid water in Monte

Carlo simulations; typically less than 5 system iterations are sufficient to achieve convergence.[33]

4. Computational Details

The computational procedure of the X-POL force field follows roughly the same approaches outlined in refs 32 and 33. However, there are several new aspects that need to be defined here. It is clear that eq 11 requires the density matrix elements, $\{P_{\eta_i \eta_j}^H(I); I = 1, \cdots, N\}$, for the auxiliary orbitals on the two boundary atoms in each residue during SCF optimizations. In the original GHO method developed for QM/MM calculations, the density is obtained by transferring the partial atomic charge on the boundary carbon from the MM force field, to the three auxiliary orbitals, plus the density of one electron.[59] In the X-POL potential, there are two hybrid orbitals from each boundary atom, and the nuclear charges are not adjusted as in QM/MM calculations; however, the major difference here is that these auxiliary orbitals are also active orbitals in the neighboring residues, which are fully optimized in SCF calculations. Consequently, the "auxiliary" densities are no longer invariant, but they are dynamically changing due to the change of molecular geometry and instantaneous charge polarizations. Furthermore, these optimized densities provide the necessary input for the auxiliary orbitals in the neighboring residues. At convergence, the chemical potential of these hybrid orbitals (active and auxiliary and vice versa) are fully equalized.

The next critical issue is to define an appropriate procedure for determining the partial charges for all other residues in eq 8 when the wave function of residue I is being optimized. A number of possibilities are available, including Mulliken population charges,[71] electrostatic potential fitted charges,[72] and the class IV (CM2) charges proposed by Cramer and Truhlar.[73–76] A good charge mapping procedure will ensure that "QM/MM" interactions be accurately determined in comparison with experimental data—through parametrization of the force field. However, special care must be taken in any charge mapping procedure such that the charge density to be used as the auxiliary density is appropriately neutralized by the atomic charges in the subunit where these "auxiliary orbitals" reside and are "active". This can be done by imposing a charge constraint if an electrostatic potential fitting procedure is used. In the present study, which is aimed at demonstrating the feasibility of the X-POL force field, we adopt the Mulliken population analysis, which properly divides the charge population between the "auxiliary orbitals" (note that they are "active orbitals" in the fragment where they are determined) and the rest of the QM subunit. Clearly, in future development of a reliable force field, both the "classical" representation of the electrostatic potential from a given wave function in terms of multipole expansions and the specific method for obtaining these multipoles shall be a primary focus of study.

Third, in the X-POL treatment, short-range exchange repulsion and long-range dispersion forces are represented by the traditional Lennard-Jones potential. Obviously, these are empirical terms that must be properly optimized against experimental and high-level ab initio results on bimolecular interactions and liquid properties including density. We have demonstrated in Monte Carlo simulations of liquid water using the X-POL approach that the van der Waals parameters in the Lennard-Jones potential can be similarly adjusted as in the development of empirical potentials, e.g., the TIP3P and TIP4P models. In the present study, we employ the corresponding parameters in the CHARMM22 force field without further modification. Note that it might be tempting to use "pure" electronic structure methods to determine the repulsive and dispersive energies; however, this would be futile in force field development if computational speed is taken into consideration.

We present an algorithm for optimization of the individual molecular wave function with the GHO boundary treatment.[59] The convergence of the entire system is achieved by an iterative SCF procedure outline below.

(a) Determine the transformation matrices $\{\mathbf{T_t}(I); I = 1, \cdots, N\}$ for the interconversion between the AOs and a set of mixed AOs and HOs, $\mathbf{C}_I^{HO} = \mathbf{T_t}(I)^{-1}\mathbf{C}_I^{AO}$, where the subscript "$\mathbf{t}$" specifies that the matrix has dimensions of $[N_I + 8] \times [N_I + 8]$. Compute the $[N_I + 4] \times [N_I + 4]$-dimensional density matrix, $\{\mathbf{P}^H(I); I = 1, \cdots, N\}$ using the active HOs for each residue.

(b) For each residue, perform SCF optimization sequentially, beginning from residue $I = 1$.

(c) For residue I, expand $\mathbf{P}^H(I)$ into full dimension by adding the auxiliary density matrix elements and transform it into the AO basis $\mathbf{P_t}^{AO}(I) = [\mathbf{T_t}(I)^{-1}]^+ \mathbf{P_t}^{HO}(I)[\mathbf{T_t}(I)^{-1}]$. Construct the full Fock matrix $\mathbf{F_t}^{AO}(I)$ in AO basis, including QM/MM interaction terms. Transform $\mathbf{F_t}^{AO}(I)$ into HO basis, $\mathbf{F_t}^{HO}(I) = [\mathbf{T_t}(I)]^+ \mathbf{F_t}^{AO}(I)[\mathbf{T_t}(I)]$. Remove the columns and rows corresponding to the auxiliary hybrid orbitals to yield the "active" Fock matrix $\mathbf{F}^{AO}(I)$, which is of $[N_I + 4] \times [N_I + 4]$-dimension. Diagonalize $\mathbf{F}^{AO}(I)$ and compute the energy and new density matrix $\mathbf{P}^H(I)$.

(d) Test convergence. If not satisfied, go to step **(c)**. If convergence is met, compute new partial charges from the optimized wave function and set the densities of the active HOs as auxiliary densities for other SCF optimizations. Incrementing I by one until $I = N$, and then, go to step **(c)**.

(e) Compute the total energy and test convergence. If not satisfied, go to step **(b)**.

We point out that the matrices transformations in step (c) are quite simple because it only involves orbitals on the boundary atoms (a total of 8 orbitals). Thus, it takes a negligible amount of computer time. As can be seen from the algorithm above, the Fock matrix construction and diagonalization are performed for each individual residue, and there are a total of N separate SCF calculations of the size of each residue in each system iteration. In our experience on the simulation of liquid water, the total number of system iterations does not increase significantly, perhaps by one or two cycles, with increased system size. Thus, the total computational time is linear scaling by $S \times N \times O(N_{max}^3)$, where S is the number of iterations in system SCF, and $O(N_{max}^3)$ is the computing efforts for the largest residue. The difference beween electronic structure calculations for a molecule of the size of $\sum_I^N N_I$ orbitals and that of N separate calculations of the size of N_{max} is obvious because the former would scale as $O([\sum_I^N N_I]^3)$ due at least to diagonalization.

Table 1. Modified Parameters for the Carbon Boundary Atom in the Present X-POL Force Field along with the Original AM1 Values and Those Used in the GHO Model[a]

parameters	AM1	GHO	X-POL
β_s	−15.715783	−5.500524	−12.85205
β_p	−7.719283	−14.666638	−5.680080
U_{ss}	−52.028658	−52.028658	−49.774256
U_{pp}	−39.614239	−38.703112	−39.573436

[a] All values are given in eV.

5. Results and Discussion

To illustrate the feasibility of the X-POL force field, we present test cases to demonstrate the procedure for optimizing parameters associated with the boundary atoms and an application to a tetrapeptide model interacting with a single water molecule. Here, we have assumed that the semiempirical AM1 model is adequate for treating the individual residues; obviously, the AM1 model itself is not satisfactory for constructing a force field for protein simulations. However, there is little doubt that they can be parametrized to accurately treat specific functional groups and interactions. The parametrization of the NDDO-based semiempirical QM model for different functional groups and atom types shall be left for future exploration.

5.1. Parametrization of Boundary Atoms. To parametrize the semiempirical force field for the boundary carbon atom and to assess its performance, we consider three model compounds: propane, 2-methylpropane (isobutene), and 2,2-dimethylpropane (neopentane). In each case, a single boundary atom is defined at the C_2 atom position, and two fragmental QM subunits are treated. We aim at the selection of a minimum number of model compounds in the parametrization process to achieve transferability by satisfying key quantum chemical requirements.[59] These include (1) the balance of electronegativity, (2) the properties of chemical bonding, and (3) the conformational potential energy profiles involving the boundary atom. Our experience in the development of the GHO methods, at the semiempirical level,[59,60] semiempirical SCC-DFTB treatment,[62] ab initio HF level,[61a,62] and DFT method,[61b] shows that if the electron-withdrawing power and the formation of the chemical bonds are adequately balanced with the QM model that the boundary atom mimics, the empirical parameters for the boundary atoms are fully transferable, just as all other standard semiempirical parameters or basis sets.

Taking the three criteria listed above into consideration, we found that we only need to modestly modify the parameters of the original AM1 Hamiltonian for carbon. We focused on the one-center core integrals U_{ss} and U_{pp} and the resonance integrals β_s and β_p, the latter of which are closely related to chemical bonding (see below). We slightly decreased these values to obtain the best overall results, but it is closer to that of the original AM1 value. Note that in the full parametrization process of the X-POL potential, the balance with all other atoms will also be consistently considered. The parameters for the boundary carbon atom are listed in Table 1 along with the standard AM1 values and those used in the GHO model.[59]

Table 2. Optimized Bond Lengths (Å) and Bond Angles (deg) Using the X-POL Potential and the Full AM1 Method

	propane		2-methylpropane		2,2-dimentylpropane	
bond	AM1	X-POL	AM1	X-POL	AM1	X-POL
C_B−C1	1.516	1.523	1.523	1.524	1.527	1.523
C_B−C3	1.516	1.526	1.523	1.524	1.527	1.526
C_B−C4(H)	1.118	1.123	1.523	1.524	1.527	1.523
C_B−C5(H)	1.118	1.120	1.121	1.118	1.527	1.519
H11−C1	1.114	1.115	1.114	1.115	1.114	1.115
H12−C1	1.114	1.113	1.114	1.114	1.114	1.114
H13−C1	1.114	1.113	1.114	1.113	1.114	1.114
H31−C3	1.114	1.115	1.114	1.114	1.114	1.114
H32−C3	1.114	1.113	1.114	1.115	1.114	1.115
H33−C3	1.114	1.113	1.114	1.114	1.114	1.114
H41−C2			1.114	1.113	1.114	1.114
H42−C4			1.114	1.115	1.114	1.115
H43−C4			1.114	1.113	1.114	1.114
H51−C5					1.114	1.115
H52−C5					1.114	1.114
H53−C5					1.114	1.114

	propane		2-methylpropane		2,2-dimentylpropane	
angle	AM1	X-POL	AM1	X-POL	AM1	X-POL
C1−C_B−C3	111.87	107.25	110.78	114.75	109.5	114.6
C1−C_B−C4			110.96	107.34	109.5	107.2
C1−C_B−C5					109.5	106.5
C3−C_B−C4			110.78	107.25	109.5	107.3
C3−C_B−C5					109.5	106.7
C4−C_B−C5					109.5	114.7

Scheme 1. Atom Numbers Assigned to the Three Alkanes Which Are Separated into Two Quantum Mechanical Fragments Across a Boundary Atom C_B
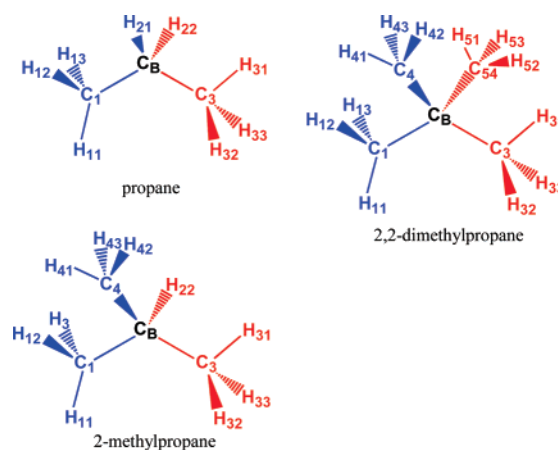
The semiempirical parameters for the resonance integrals, β_s and β_p, are most directly responsible for chemical bonding and molecular geometry. The optimized parameters in the table show that the X-POL potential has very similar values compared with the original AM1 parameters. This is in contrast to the GHO method,[59] which does not have the double self-consistent field treatment to optimize the auxiliary hybrid orbital densities. The bond lengths and selected bond angles of propane, 2-methylpropane, and 2,2-dimethylpropane, optimized using the X-POL potential and the AM1 method, are given in Table 2 (see Scheme 1 for atom assignment). The key parameters to be examined are the bond

Table 3. Mulliken Population Charges (au) Obtained Using the X-POL Potential and the Full AM1 Method. Values in Parentheses Are Sums Over Hydrogens

bond	propane		2-methylpropane		2,2-dimentylpropane	
	AM1	X-POL	AM1	X-POL	AM1	X-POL
C_B	−0.160	−0.159	−0.111	−0.107	−0.060	−0.054
C1	−0.210 (0.004)	−0.230 (0.003)	−0.206 (0.010)	−0.223 (0.012)	−0.202 (0.015)	−0.230 (0.013)
H11	0.071	0.072	0.072	0.072	0.072	0.072
H12	0.071	0.084	0.07203	0.085	0.072	0.086
H13	0.071	0.077	0.072	0.078	0.072	0.086
C3	−0.210 (0.004)	−0.235 (−0.003)	−0.206 (0.010)	−0.223 (−0.001)	−0.202 (0.015)	−0.230 (0.007)
H31	0.071	0.072	0.072	0.085	0.072	0.085
H32	0.071	0.083	0.072	0.073	0.072	0.072
H33	0.071	0.077	0.072	0.077	0.072	0.086
C4 (H)	0.076	0.076	−0.206 (0.010)	−0.242 (0.012)	−0.202 (0.015)	−0.236 (0.013)
H41			0.072	0.084	0.072	0.085
H42			0.072	0.073	0.072	0.073
H43			0.072	0.084	0.072	0.085
C5 (H)	0.076	0.083	0.081	0.084	−0.202 (0.015)	−0.225 (0.020)
H51					0.072	0.072
H52					0.072	0.086
H53					0.072	0.086

lengths and bond angles associated with the boundary atom, C_B, which is placed at the C2 position in all three cases. The present comparison is best made with the values optimized using the AM1 model, rather than the experimental or high-level ab initio results, because the main purpose is to evaluate the possibility to parametrize the boundary atom to reproduce the results from the QM model used to describe the QM subunits. For the nine C_B−C bonds and three C_B−H bonds in these three model compounds, the average unsigned error is 0.004 and 0.003 Å, respectively. The agreement with the AM1 results is good. Bond angles see somewhat greater variations mainly because the way that the hybrid orbitals are defined. In the original GHO approach, the hybrid orbital pointing toward the QM fragment is defined based on the local (instantaneous) geometry of the other (MM) three bonds connected to the boundary atom.[59] The remaining three auxiliary hybrid orbitals are created by using Schmidt orthorgonalization and equal hybridization.[60] In the present application, we have adopted the same definition and partitioning scheme for the hybrid orbitals, but, of course, all hybrid orbitals should be treated equally based on the respective local geometry followed by a Lowdin-type orthorgonalization. The latter approach would resolve the slight imbalance caused by the hybridization method. Nevertheless, the optimized bond angles are still reasonable for the present test purposes.

On the other hand, the fundamental criterion necessary to ensure transferability of these atomic parameters for boundary atoms is the balance of the electron-withdrawing power of the boundary atom with that of the QM model that it mimics (AM1 in the present case). Therefore, the boundary atom must have the same electronegativity as that of an "AM1 carbon" atom so that there is no charge transfer between two identical groups. The most relevant parameters in the semiempirical theory are the one-center U_{ss} and U_{pp} terms, which are optimized in connection with the resonance integral parameters (as they are not independent in energy calculations). We found that it is possible to achieve this

goal by only optimizing these four parameters listed in Table 1. This is illustrated by the computed Mulliken population charges for the three alkane model compounds. It perhaps should be emphasized here that the Mulliken population[71] is in fact the best charge analysis to examine the balance of the electron-withdrawing abilities of different elements or between atoms of the same type, but they are treated differently (AM1 vs GHO).

Ideally, there is no net charge transfer between two neighboring groups across the boundary atom in the X-POL potential, although a slight variation is inevitable since the boundary atom is, after all, an approximation to the original QM method. Propane is used as the primary target in the parametrization, and the goal is to have as little charge transfer as possible between two fragments via C_B. Table 3 shows that the sum of the total charges of the two QM fragments are nearly the same (0.080 au) both from AM1 and the X-POL calculations, and the partial charge on the boundary carbon only differs by 0.001 au (Scheme 1). To remove the effect of the hydrogen atoms on the C_B atom, which have not been reparameterized, we examine the second symmetric system, neopentane. The unrestrained AM1 calculation yields a total net charge of 0.030 au for two methyl groups, which may be compared to values of 0.026 and 0.027 au from the X-POL potential. Importantly, the two fragments are reasonably balanced. The small difference among the individual methyl groups is again due to the definition of the hybrid orbitals for the boundary atom, which are not completely equivalent. For isobutene, the difference between the two QM fragments is 0.060 au more positive for one methyl group and an H_B than two methyl groups from full AM1 calculations, whereas the difference is 0.059 au. The agreement between AM1 and X-POL partial charges on the C_B atom is also good, which shows the absolute amount of charge imbalance between the boundary atom and the standard carbon.

The torsional potential energy profiles for isobutene about a C−C dihedral are illustrated in Figure 3. The torsional
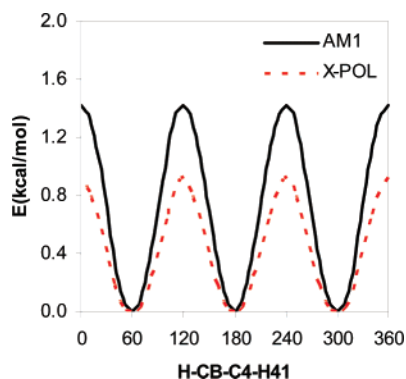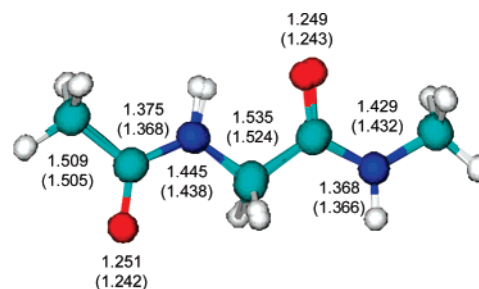
Figure 3. Torsional potential energy profiles for 2-methyl-propane (isobutene) about the H22−CB−C4−H41 dihedral angle from AM1 and X-POL optimizations. Energies are in kcal/mol and dihedral angles are in degrees.

Figure 4. Superposition of the optimized structures using the AM1 and the X-POL potentials for glycine dipeptide. Bond lengths in angstroms are given for the X-POL potential and for the AM1 model in parentheses.

Table 4. Mulliken Population Charges (au) for Glycine Obtained Using the X-POL Potential and the Full AM1 Method

atom	AM1	X-POL $C_\alpha-C$	X-POL $C_\alpha-N$
N	−0.349	−0.350	−0.348
HT1	0.157	0.128	0.126
HT2	0.151	0.180	0.160
$H_\alpha 1$	0.122	0.108	0.120
C_α	−0.047	−0.056	−0.067
$H_\alpha 2$	0.085	0.126	0.120
C	0.302	0.266	0.283
OT1	−0.342	−0.374	−0.355
OT2	−0.314	−0.282	−0.291
HO2	0.236	0.254	0.252

Table 5. Mulliken Population Charges (au) for Glycine Dipeptide Obtained Using the X-POL Potential and the Full AM1 Method

atom/group	AM1	X-POL $C_\alpha-C$	X-POL $C_\alpha-N$
CH_3	0.060	0.060	0.058
CO	−0.067	−0.091	−0.094
NH	−0.130	−0.132	−0.142
$H_\alpha 1$	0.111	0.096	0.108
C_B	−0.039	−0.002	−0.013
$H_\alpha 2$	0.111	0.128	0.121
CO	−0.085	−0.133	−0.111
NH	−0.147	−0.123	−0.124
CH_3	0.185	0.195	0.197

AM1 values, while geometrical parameters are presented in Figure 4.

Overall the partial charges show reasonable transferability. The amino group has a total net partial charge of −0.041 au from AM1, whereas it is −0.042 and −0.062 au from the X-POL potential when the hybrid orbitals are defined based on the $C_\alpha-C$ vector and the $C_\alpha-N$ vector, respectively. For the carboxyl group, the total net charges are −0.136, −0.111, and −0.118 au from AM1, X-POL ($C_\alpha-C$ bond), and X-POL ($C_\alpha-N$ bond), respectively. The difference is only about 0.02 au for such a strong electron-withdrawing group. Overall, there is little charge transfer between the two QM fragments for both partition schemes in comparison with the AM1 partial charges.

The optimized structure of glycine dipeptide for the extended conformation is illustrated in Figure 4 along with some selected bond lengths from the X-POL and AM1 potentials. Overall, the structural agreement is excellent with a root-mean-square difference of 0.007 Å for bonds shown in Figure 4. The combined group charges are given in Table 5; the trend of charge polarization is adequately retained in the X-POL potential, and the charge delocalization across the boundary atom is also good.

energy from the X-POL potential has contributions from "pure" QM, QM/MM, and "pure" MM (van der Waals) terms, and the computed barrier height is about 0.5 kcal/mol lower than the full AM1 energy. This trend has been observed in the GHO boundary approach,[59] and this difference is easily corrected by adjusting the semiempirical method or by including a classical torsional term. The latter scenario is probably a simple choice for constructing an empirical force field, and this would be the only internal bonding terms required in the X-POL potential.

5.2. Glycine and Glycine-Dipeptide. To examine the possibility that the X-POL potential can be applied to proteins, we optimized the structures of glycine and glycine dipeptide, which are compared with the full original AM1 calculations. Note that the main goal here is to show that the boundary parameters optimized above are transferable to polypeptides and that the development of the X-POL potential will involve full optimization of the QM model itself for different functional groups. Listed in Table 4 are computed partial atomic charges for glycine when the boundary atom is placed on the C_α atom with the first hybrid orbital pointed either toward the carbonyl carbon ($C_\alpha-C$) or toward the amino nitrogen ($C_\alpha-N$), as the current definition of the hybrid orbitals still does not yield exactly equivalent hybridizations. The results are compared with the

6. Summary

An electronic structure-based polarization potential, which is called the X-POL potential, has been described for the purpose of constructing an empirical force field for modeling polypeptides. The X-POL potential takes an entirely different

Design of a Next Generation Force Field

J. Chem. Theory Comput., Vol. 3, No. 6, 2007 **1899**

philosophical approach toward the development of a force field and the treatment of electronic polarization and charge delocalization. The internal, bonded interactions are fully represented by an electronic structure theory augmented with some empirical torsional terms. Nonbonded interactions are modeled by an iterative, combined quantum mechanical and molecular mechanical method, in which the molecular mechanical partial charges are derived from the molecular wave functions of the individual fragments. In this paper, the X-POL potential is illustrated by making use of the neglect of diatomic differential overlap (NDDO) approximation and the AM1 model as the quantum mechanical method, without further parametrization for specific functional groups. The main purpose of this study is to demonstrate the feasibility of such an electronic structure force field and to develop a practical and well-defined method for separating a polypeptide chain into peptide units. The boundary is treated following the ideas of the generalized hybrid orbital (GHO) technique developed for the treatment of QM and MM boundaries and extended to bridge two QM regions in the X-POL potential. The parametrization procedure and philosophy for the boundary treatment between QM fragments in the X-POL potential is documented and tested by a number of simple compounds.

The X-POL model presented here is an empirical *force field*, although it is based on quantum mechanical formalisms. If one finds that a particular QM model used or the approximations made are inadequate to treat certain properties, for example, the torsional potential energy profile about a single bond rotation, one can include a purely empirical energy term such as the sine and cosine function series in the current force fields. Although this might be deemed ad hoc, the method is nevertheless systematic in that one can always seek for a better, more accurate QM representation of the individual residues such that these empirical functional terms can be eliminated.

In two forthcoming papers, we describe the analytical energy gradient techniques for the X-POL potential and an application to a solvated protein. The exact treatment and construction of individual force fields in the future may differ from the method presented here, but the general direction seems to be clear. We envision that the next generation of force fields for biomolecular polymer simulations will be developed based on electronic structure theory, which is one way to properly define and treat many-body polarization and charge delocalization effects.

References

(1) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.

(2) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584.

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

(4) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(5) 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.

(6) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Eur. Biophys. J.* **2005**, *34*, 273.

(7) Maple, J. R.; Hwang, M. J.; Jalkanen, K. J.; Stockfisch, T. P.; Hagler, A. T. *J. Comput. Chem.* **1998**, *19*, 430.

(8) Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497.

(9) Dinur, U.; Hagler, A. T. *Rev. Comput. Chem.* **1991**, *2*, 99.

(10) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.

(11) Warshel, A. *J. Chem. Phys.* **1994**, *101*, 6141.

(12) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

(13) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208.

(14) Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.

(15) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689.

(16) (a) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141. (b) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741.

(17) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

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

(19) Patel, S.; Brooks, C. L., III. *Mol. Simul.* **2006**, *32*, 231.

(20) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185.

(21) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587.

(22) Lee, T. S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 7549.

(23) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1998**, *120*, 5593.

(24) Van der Vaart, A.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1999**, *121*, 9182.

(25) Mo, Y.; Gao, J. *J. Phys. l Chem. B* **2006**, *110*, 2976.

(26) (a) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049. (b) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.

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

(28) Gao, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. In *Rev. Comput. Chem.*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 7, pp 119.

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

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

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

(32) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.

(33) Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346.

(34) Wierzchowski, S. J.; Kofke, D. A.; Gao, J. *J. Chem. Phys.* **2003**, *119*, 7365.

(35) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744.

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

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

(38) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.

(39) Rothlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem.* **2000**, *5*, 236.

(40) Tuckerman, M. E.; Marx, D.; Parrinello, M. *Nature* **2002**, *417*, 925.

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

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

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

(44) Zerner, M. C. *Rev. Comput. Chem.* **1991**, *2*, 313.

(45) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374.

(46) Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492.

(47) Gao, J. *J. Comput. Chem.* **1997**, *18*, 1062.

(48) Gao, J.; Byun, K. *Theor. Chem. Acc.* **1997**, *96*, 151.

(49) Lin, Y.-l.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1484.

(50) Poulsen, T. D.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys.* **2002**, *116*, 3730.

(51) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050.

(52) Wesolowski, T.; Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1996**, *100*, 15444.

(53) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

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

(55) Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175.

(56) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(57) Dapprich, S.; Komiromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *Theochem* **1999**, *461−462*, 1.

(58) Field, M. J. *Mol. Phys.* **1997**, *91*, 835.

(59) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.

(60) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336.

(61) (a) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632. (b) Pu, J.; Gao, J.; Truhlar, D. G. *Chem. Phys. Chem.* **2005**, *6*, 1853.

(62) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 5454.

(63) (a) Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610. (b) Thery, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269. (c) Assfeld, X.; Ferré, N.; Rivail, J.-L. *ACS Symp. Ser.;* Gao, J., Thompson, M. A., Eds.; 1998; Vol. 712, p 234.

(64) Antes, I.; Thiel, W. *ACS Symp. Ser.;* Gao, J., Thompson, M. A., Eds.; 1998; Vol. 712, p 50.

(65) Calzaferri, G.; Forss, L.; Kamber, I. *J. Phys. Chem.* **1989**, *93*, 5366.

(66) Carbo, R.; Fornos, J. M.; Hernandez, J. A.; Sanz, F. *Int. J. Quantum Chem.* **1977**, *11*, 271.

(67) Anderson, A. B.; Hoffmann, R. *J. Chem. Phys.* **1974**, *60*, 4271.

(68) Elstner, M.; Porezag, D.; Juugnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Sukai, S.; Seifect, G. *Phys. Rev. B* **1998**, *58*, 7260.

(69) Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614.

(70) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907.

(71) Mulliken, R. S. *J. Chem. Phys.* **1964**, *61*, 20.

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

(73) Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.

(74) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

(75) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291.

(76) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503.