ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields

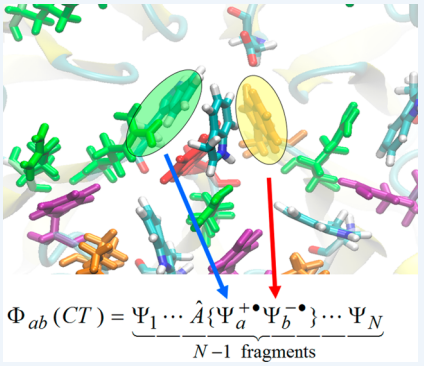
Jiali Gao,*[,†,‡] Donald G. Truhlar,[‡] Yingjie Wang,[‡] Michael J. M. Mazack,[‡] Patrick Löffler,[‡] Makenzie R. Provorse,[‡] and Pavel Rehak[‡]

[†]Theoretical Chemistry Institute, State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun, Jilin Province 130028, People's Republic of China

[‡]Department of Chemistry and Supercomputing Institute University of Minnesota, Minneapolis, Minnesota 55455, United States

Supporting Information

CONSPECTUS: Molecular mechanical force fields have been successfully used to model condensed-phase and biological systems for a half century. By means of careful parametrization, such classical force fields can be used to provide useful interpretations of experimental findings and predictions of certain properties. Yet, there is a need to further improve computational accuracy for the quantitative prediction of biomolecular interactions and to model properties that depend on the wave functions and not just the energy terms. A new strategy called explicit polarization (X-Pol) has been developed to construct the potential energy surface and wave functions for macromolecular and liquid-phase simulations on the basis of quantum mechanics rather than only using quantum mechanical results to fit analytic force fields. In this spirit, this approach is called a quantum mechanical force field (QMFF).



X-Pol is a general fragment method for electronic structure calculations based on the partition of a condensed-phase or macromolecular system into subsystems ("fragments") to achieve computational efficiency. Here, intrafragment energy and the mutual electronic polarization of interfragment interactions are treated explicitly using quantum mechanics. X-Pol can be used as a general, multilevel electronic structure model for macromolecular systems, and it can also serve as a new-generation force field. As a quantum chemical model, a variational many-body (VMB) expansion approach is used to systematically improve interfragment interactions, including exchange repulsion, charge delocalization, dispersion, and other correlation energies. As a quantum mechanical force field, these energy terms are approximated by empirical functions in the spirit of conventional molecular mechanics. This Account first reviews the formulation of X-Pol, in the full variationally correct version, in the faster embedded version, and with systematic many-body improvements. We discuss illustrative examples involving water clusters (which show the power of two-body corrections), ethylmethylimidazolium acetate ionic liquids (which reveal that the amount of charge transfer between anion and cation is much smaller than what has been assumed in some classical simulations), and a solvated protein in aqueous solution (which shows that the average charge distribution of carbonyl groups along the polypeptide chain depends strongly on their position in the sequence, whereas they are fixed in most classical force fields). The development of QMFFs also offers an opportunity to extend the accuracy of biochemical simulations to areas where classical force fields are often insufficient, especially in the areas of spectroscopy, reactivity, and enzyme catalysis.

1. INTRODUCTION

Molecular mechanical (MM) force fields are widely used in computer simulations of macromolecular systems, including proteins and nucleic acids in aqueous solution.[1,2] Although these models are computationally efficient for applications to large systems, there are also a number of well-known limitations that are not easily resolved in the MM context.[3,4] There is a great need to further improve the accuracy to achieve quantitative prediction of biomolecular interactions such as ligand binding, electron and energy transfer, and enzymatic reactions. It is timely to ask what type of potential energy functions will be used in the future for biomolecular simulations. One possibility is to continue improving the current MM approach, including classical polarization terms and charge transfer in the force fields. The second is to develop a new theoretical framework with the capability of including quantum mechanical effects explicitly to model intermolecular interactions.[5−9]

Quantum mechanics, in principle, can provide both reactive and nonreactive potential energy surfaces, including electrostatics, electron correlation, polarization, and charge transfer. Yet, it is limited by the computational cost, which increases rapidly with the size of the system. To overcome this limitation, a variety of fragment-based electronic structure methods,[10]

Special Issue: Beyond QM/MM: Fragment Quantum Mechanical Methods

Received: June 9, 2014
Published: August 6, 2014

   dx.doi.org/10.1021/ar5002186 | Acc. Chem. Res. 2014, 47, 2837−2845

many of which are illustrated in this Special Issue, have been developed to increase the computational speed for large molecular systems. For example, Gordon and co-workers developed the effective fragment potential method,[11] York and co-workers employed a version of the divide-and-conquer strategy for linear-scaling quantum-mechanical (QM) representation of a macromolecular system as a force field,[8,9,12] and Herbert and co-workers discussed the issue of error propagation in many-body expansion theory.[13] A key issue is the requirement for a suitable level of electronic structure theory for monomers.[13,14] For brevity, we shall not make an extensive list of fragment methods, but the introduction of a recent paper may be consulted for ∼70 relevant references.[31] This Account presents one especially promising fragment method, namely, the explicit-polarization method (X-Pol),[5,6,14−17] for the construction of a next-generation force field on the basis of quantum mechanics.[18]

The X-Pol model is a fragment-based QM method,[5,6,17] in which the entire system is divided into molecular subunits, which can consist of one or more molecules, ions, ligands, cofactors, or amino acid residues. The key assumption in the X-Pol method is that the wave function of the entire system is approximated as a Hartree product of the antisymmetric wave functions of individual fragments.[5] Then, the optimization of the total wave function can be reduced to the optimization of each fragment embedded in and polarized by the rest of the system; this reduces the computational cost to nearly linear scaling with respect to the number of fragments.[19] Clearly, variational optimization of the mutual dependence of the fragmental wave functions is critical to the success of this method.[16,17] As a force field, the energy of each fragment is determined by the electronic structure method used, whereas intermolecular interactions are modeled through electrostatic embedding.[15] The short-range exchange-repulsion interactions, charge transfer, and long-range dispersion interactions and correlation energy between different fragments can be modeled empirically as in MM.[5,6,15,20] Alternatively, these energy contributions can be modeled by density-dependent functionals, by Hartree−Fock (HF) exchange, or by making use of many-body expansion corrections.[10,14,38]

In the following, we describe the theoretical approach of X-Pol and illustrate computational results that demonstrate the feasibility and accuracy of macromolecular simulations employing the X-Pol quantum mechanics force field (QMFF).

2. THEORETICAL BACKGROUND

In X-Pol, a system is partitioned into molecular fragments,[5,6,15] which may be called monomers. For solutions with small solute molecules, a fragment can be a single solute or solvent molecule.[15,18,21] For a polypeptide chain, the peptide unit as defined by the IUPAC nomenclature (rather than the conventional residue)[22] is assigned to be the smallest fragment that contains the atoms $-C_\alpha^I R^I CO-NHC_\alpha^{I+1}H-$ in the conventional residues I and I + 1, where each C_α atom is shared by two adjacent peptide units (Figure 1).[6,23] Several peptide units can be combined into the same fragment, if desired, which can be useful for modeling systems containing disulfide bonds. The connection approach is an extension of the generalized hybrid orbital (GHO) method,[24] originally developed for combined QM/MM applications.[25] Unlike schemes that involve capping by a hydrogen atom,[26] the GHO method does not alter the number of degrees of freedom or the electrostatic interactions between neighboring fragments.

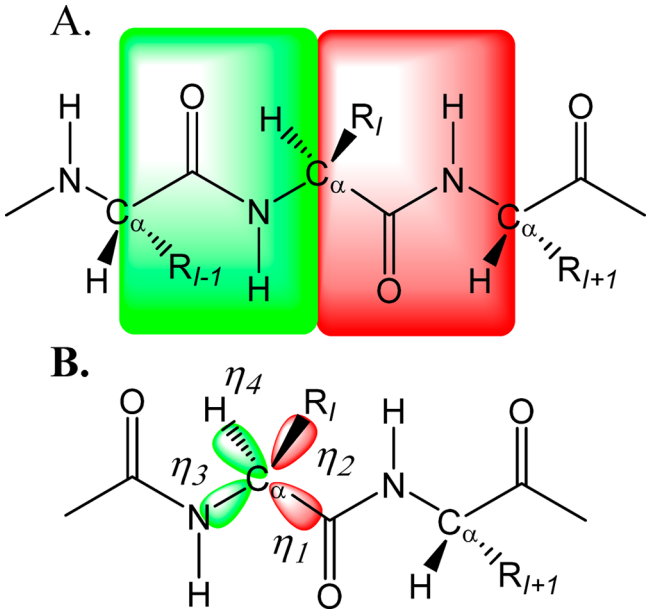


Figure 1. Schematic depiction of the division of a polypeptide chain into peptide units (A). Two fragments are highlighted in green and red, respectively, corresponding to residues I − 1 and I. The C_α boundary atom connecting these two peptide units is shown in panel B, and its four hybrid orbitals are equally partitioned into the two neighboring fragments.

Thus, the hybrid orbital approach provides a seamless transition from one fragment to the next across chemical bonds, and a buffering scheme was used to accelerate the convergence.[27]

The X-Pol model is derived from a conventional electronic structure method by a hierarchy of three approximations.[5,6,17] First, the wave function of the entire system is approximated as a Hartree product of the antisymmetric wave functions of the individual fragments,

$$\Phi = \prod_A^N \Psi_A \tag{1}$$

where N is the number of fragments. The effective Hamiltonian of the system is given by

$$\hat{H} = \sum_A^N \hat{H}_A^o + \frac{1}{2}\sum_A^N \sum_{B\neq A}^N (\hat{H}_A^{int}[V_E^B] + \Delta E_{AB}^{XCD}) \tag{2}$$

where the first term sums over the Hamiltonians of all isolated fragments $\{\hat{H}_A^o\}$, and the double summation accounts for pairwise interactions among all the fragments; $\hat{H}_A^{int}[V_E^B]$ represents electrostatic interactions between fragments A and B with V_E^B being the potential due to B, and the final term, ΔE_{AB}^{XCD}, specifies exchange-repulsion (X), charge delocalization (C), and dispersion and other interfragment correlation (D) energy contributions.

A general approach for representing the external potential, V_E^B, due to the charge density of fragment B is to use a multicenter multipole expansion,[5,6,17,28] of which the simplest form is to limit the expansion to the monopole terms, so the result only depends on the partial atomic charges.[15] The use of partial atomic charges to approximate V_E^B is particularly convenient for constructing the effective Hamiltonian of eq 2, and this is the strategy that has been adopted in the X-Pol

dx.doi.org/10.1021/ar5002186 | Acc. Chem. Res. 2014, 47, 2837−2845

method.[5,6,15−18,21] This strategy has also been used in other studies.[9,13,29−32]

The total energy of the system in the X-Pol method is given by

$$E_{\text{X-Pol}} = \langle \Phi | \hat{H} | \Phi \rangle \tag{3}$$

There are two ways of constructing the Fock matrix for optimizing the fragmental wave functions, and they distinguish the variational X-Pol method from other fragment-based methods.

2.1. Variational X-Pol

With the use of partial atomic charges to approximate V_{E}^{B}, the Fock operator for a fragment, A, is derived variationally to yield[16,17,27]

$$\mathbf{F}_{\mu\nu}^{A,\text{Xpol}} = \mathbf{F}_{\mu\nu}^{A,\text{o}} - \frac{1}{2}\sum_{B\neq A}\sum_{b\in B} q_b^B (\mathbf{I}_b^B)_{\mu\nu}^A + \frac{1}{2}\sum_{a\in A} X_a^A (\mathbf{\Lambda}_a^A)_{\mu\nu} \tag{4}$$

where $\mathbf{F}_{\mu\nu}^{A,\text{o}}$ is the Fock matrix element for the Hamiltonian of isolated fragment A, q_b^B is the point charge on atom b of fragment B, $\mathbf{I}_b^B$ is the matrix of the one-electron integrals of the embedding potential due to fragment B, and the last term represents the response of the charge density of fragment B due to variational optimization of the wave function of fragment A (X_a^A is the derivative of the energy with respect to the atomic charge on a, and $\mathbf{\Lambda}_a^A$ is the charge derivative with respect to the density, see ref 33).

2.2. Embedded X-Pol

If each fragment is considered to be embedded in the static field of the rest of the system, one can construct a Fock operator for fragment A heuristically,[5,11,15,34] akin to a combined QM/MM approach;[25] this yields

$$\mathbf{F}_{\mu\nu}^{A,\text{eX}} = \mathbf{F}_{\mu\nu}^{A,\text{o}} - \sum_{B\neq A}\sum_{b\in B} q_b^B (\mathbf{I}_b^B)_{\mu\nu}^A \tag{5}$$

In this charge-embedding approach, the mutual polarization among all fragments in the system is achieved by iteratively updating the partial atomic charges. However, eq 5 is not variational, and it does not include the response of the charge density to the changes of other charges. To emphasize this difference and to emphasize that the full X-Pol algorithm is a variational method, in the rest of this Account, the method in section 2.1 will be called variational X-Pol, and the subsequent many-body expansion corrections (to be introduced below) will be called VMB2 and VMB3. The embedding approach (section 2.2) is simply called embedded X-Pol, and the many-body corrections to it will be called EMB2 and EMB3. When we just say X-Pol, MB2, or MB3, it applies to either.

The variational X-Pol method has the advantage over embedded X-Pol of allowing the computation of analytic gradients for efficient geometry optimization and dynamics simulations.[16,17,27] Furthermore, the total energy obtained from the variational procedure is necessarily lower than that from the charge-embedding scheme. Consequently, it is expected that the use of the variational X-Pol energy as the monomer energy reference state in many-body energy expansion should lead to smaller corrections than other alternatives.[14] Although it is possible to obtain analytic gradients for the nonvariational, charge-embedding approaches,[35] it generally involves solution of coupled-perturbed self-consistent field equations. Sometimes the response terms in fragment-based methods have been neglected.[36]

2.3. Many-Body Improvements

The errors in the procedures presented so far are sometimes unacceptably large. In such cases, we can consider those procedures to constitute the one-body term in a many-body expansion and systematically improve the results by including higher-order terms in the expansion.

Two alternatives can be followed, which correspond to the use of X-Pol as a multilevel quantum chemical method,[14,36,37] as exemplified by the examples to be presented in section 3.1, or as a quantum mechanical force field that introduces empirical energy terms,[5,6,18] as in section 3.2. The Hartree product wave function in eq 1 implies that interfragment charge delocalization and exchange-repulsion interactions arising from the Pauli exclusion principle are neglected.[38,39] However, these interactions as well as interfragment dispersion interactions make critical contributions to intermolecular interactions. Together, they are included in a term called ΔE_{XCD}, and they must be properly accounted for.[38−40] A brute force approach is to employ variational many-body expansion (VMB) theory to make two-body, three-body, and higher order corrections,[10,14] but nonvariational approaches are also possible.[11,30,34] When two- and three-body terms are included, we have, respectively, the corresponding VMB2 and VMB3 corrections,

$$\Delta E_{\text{XCD}}^{(2)} \approx \Delta E_2^{\text{VMB}} = \sum_{I<J}^{N} (E_{IJ} - E_{\text{X-Pol}}) \tag{6}$$

$$\Delta E_{\text{XCD}}^{(3)} \approx \Delta E_3^{\text{VMB}} = \sum_{I<J<K}^{N} (E_{IJK} - E_{\text{X-Pol}} - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK}) \tag{7}$$

In eqs 6 and 7, E_{IJ} and E_{IJK} are, respectively, the X-Pol energies when two and three monomers are grouped into a single superfragment. Thus, the difference in parentheses in eq 6 gives the two-body correction energy, ΔE_{IJ}, accounting for exchange and charge delocalization effects at the Hartree−Fock level plus dispersion-correlation contributions when correlated methods are used. Typically, the dimer and trimer fragments are significantly smaller than the full system, making the computation costs for a single oligomer negligible compared with the cost for a single calculation of the entire system. However, the number of trimer and higher-order terms increases rapidly with the number of fragments, rendering these terms impractical. Thus, in using this approach, it is critical to define a reference state for the monomer energy such that terms of order higher than 2 are negligible.[14,37]

When the X-Pol method is used as a theoretical framework to develop force fields for condensed-phase and macromolecular systems, a simpler, empirical approach can be adopted, such as Lennard-Jones or Buckingham potentials (as used in molecular mechanics)[2,3,8,15,17,20] or perturbation theory[41] to estimate the dimer $\Delta E_{AB}^{\text{XCD}}$ terms.

3. ILLUSTRATIVE EXAMPLES

3.1. Multilevel X-Pol as a Quantum Chemical Model for Macromolecules

The X-Pol method can be used with different electronic structure representations for different fragments. This provides

dx.doi.org/10.1021/ar5002186 | Acc. Chem. Res. 2014, 47, 2837−2845

a general, multilevel QM/QM treatment for large systems, where the region of most interest could be modeled by a high-level electronic structure theory, embedded in an environment modeled by a lower-level representation.[33] To illustrate this possibility, a number of formulations of different electronic structure models have been used to describe the bimolecular complexes between CH_3CO_2H and H_2O and between a Zundel ion $H_5O_2^+$ and four water molecules.[33] In multilevel X-Pol calculations with different QM models, the interaction between different fragments is coupled by the electrostatic potential, V_E^B, in eq 2. Two different charge models, namely, Mulliken population analysis (MPA) and electrostatic potential (ESP) charge-fitting, have been used to construct the charge-embedding Fock matrix (eq 5), whereas only the MPA charges were used in the variational X-Pol (eq 4) model.

The variational and embedded X-Pol interaction energies between $H_5O_2^+$ and $(H_2O)_4$ are listed in Table 1, and they tend to overestimate the binding interactions since interfragment exchange-repulsion interactions are neglected. With the inclusion of the ΔE_{XCD} energy, derived from an energy decomposition analysis[40] using CCSD/MG3S and HF/MG3S results, the computed binding energies are significantly improved. The variational CCSD/MG3S:M06/MG3S combination yielded reasonable agreement (−68.0 kcal/mol) with the binding energy obtained from a higher-level CCSD(T)/MG3S calculation (−69.7 kcal/mol). The errors from the other combinations in the table range from 2.3 to 8.9 kcal/mol. This demonstrates that it is important to develop a systematic procedure to improve the accuracy of X-Pol (both variational and embedding models) beyond the monomer approximation.

Table 1. Computed Electrostatic Interaction Energies, ΔE_{elec} (kcal/mol), between $H_5O_2^+$ (A) and $(H_2O)_4$ (B) Using Multilevel X-Pol with the Charge-Embedded and Variational Interaction Hamiltonians[a]

		charge-embedding		variational		
A	B	ESP	MPA	MPA	ΔE_{XCD}	ΔE_b
M06	M06	−89.1	−87.5	−91.0	15.9	−75.1
M06	B3LYP	−87.7	−85.2	−88.1	15.9	−72.2
M06	HF	−92.0	−91.7	−94.5	15.9	−78.6
MP2	HF	−92.9	−92.7	−94.4	15.9	−78.5
CCSD	M06	−89.5	−88.0	−83.9	15.9	−68.0

[a]The 6-31G(d) basis set was used in all calculations with M06/MG3S optimized monomer and dimer geometries, and ΔE_{XCD} was estimated using HF and CCSD/MG3S energies.

The errors due to the Hartree-product approximation for the molecular wave function in X-Pol can be systematically corrected using variational[10,14] and embedding[5,34,42−45] many-body expansions (VMB*n* and EMB*n*). Table 2 displays the total interaction energies of four local-minimum-energy structures for the water hexamer, determined up to the third order VMB correction, as calculated by the Hartree−Fock (HF) and recently introduced polarized molecular orbital (PMO) methods.[18] Although there are significant errors at the monomer level due to neglecting the ΔE_{XCD} contributions, the binding energies are significantly improved with the dimer correction (VMB2),[14] and the computed results at the VMB3 level are in quantitative agreement with those from the corresponding full QM calculations.

The VMB results are in systematically better agreement with the binding energies from full QM calculations than are the results of the corresponding charge-embedding approach. The difference is particularly striking at the two-body correction level, with average errors of 2.2 and 1.5 kcal/mol from VMB2@HF and VMB2@PMO, as compared with 4.2 and 6.1 kcal/mol for EMB2@HF and EMB2@PMO, respectively. This emphasizes the critical role of defining a good monomer energy in MB expansion theory, and the variational X-Pol energy leads to better convergence in correction terms. Including three-body energies, the errors are of chemical accuracy.

Although results for the small water clusters are encouraging, we found that significant errors exist for larger water clusters (Table 3). We examined a system containing 65 molecules, taken from a configuration in liquid simulations by keeping water molecules within 7.5 Å of a selected monomer center. We used HF, MP2, B3LYP and B3LYP-D, and PMO electronic structure methods to describe the monomers and dimers, all with the 6-31G(d) basis set (except PMO, which has a method-specific basis set). The absolute binding energies are not converged in these calculations; one may note in passing that the performance of the semiempirical PMO model[18] for water is exceptionally good, but the main purpose here is to test the performance of many-body expansion theory. Among the four ab initio and density functional models, the average unsigned error using the variational model at the VMB2 level is about 60 kcal/mol, or about 1 kcal/mol per monomer. We see that the error is small employing the EMB2 approach with EX-Pol background charges; in particular, it is about 11 kcal/mol for the complex or 0.17 kcal/mol per monomer. For comparison, a fragment molecular orbital calculation (in particular by the FMO2 procedure, which is very similar to the MB2 approach) for a 64-water cluster yielded an error of 0.003 au (1.9 kcal/mol) per monomer by employing MP2.[46] The origin of

Table 2. Computed Interaction Energies for Water Hexamer Structures from Unfragmented Quantum Mechanical Calculations, The Variational Many-Body Expansion (VMB), and Embedded Many-Body Expansion (EMB) Models at the HF/6-311G(d,p) and Polarized Molecular Orbital (PMO) Levels

	book		cage		cyclic		prism	
	HF	PMO	HF	PMO	HF	PMO	HF	PMO
Full QM	−45.1	−44.9	−45.9	−47.6	−44.3	−41.8	−47.1	−46.9
VX-Pol	−53.9	−26.8	−49.9	−25.0	−58.0	−28.7	−49.1	−24.3
VMB2	−42.5	−43.6	−44.9	−45.7	−39.8	−40.0	−46.2	−45.9
VMB3	−44.4	−44.8	−44.9	−46.5	−43.6	−41.6	−46.0	−46.8
EX-Pol	−39.2	−22.5	−37.2	−21.2	−41.2	−24.1	−36.8	−20.7
EMB2	−40.3	−39.0	−43.2	−41.1	−37.5	−35.3	−44.7	−41.3
EMB3	−44.4	−44.2	−44.9	−46.1	−43.3	−40.9	−45.9	−46.3

dx.doi.org/10.1021/ar5002186 | Acc. Chem. Res. 2014, 47, 2837−2845

Table 3. Computed Interaction Energies for a $(H_2O)_{65}$ Water Cluster Using Variational and Embedding X-Pol and Two-Body Corrections[a]

	Full QM	VX-Pol	VMB2	EX-Pol	EMB2
HF	−645.5	−1008.7	−691.2	−942.4	−638.8
MP2	−975.6	−915.3	−917.3	−877.3	*b*
B3LYP	−887.1	−861.5	−953.7	−797.6	−874.4
B3LYP-D	−1042.7	−861.5	−1109.4	−797.6	−1030.1
PMO	−735.5	−432.6	−720.9 (−734.7)[c]	−344.8	−602.6 (−722.3)[c]

[a]The 6-31G(d) basis set was used in ab initio and DFT calculations. [b]Not calculated. [c]Entries in parentheses are by VMB3 and EMB3.

relatively large errors in VMB is due to charge penetration effects in X-Pol optimization. While this is not significant with a minimal basis such as that employed in semiempirical methods, it can become unphysically large when extended basis sets are used.[47] In some cases, an explicit exchange repulsion potential may be needed to prevent these effects.

Charge penetration effects are much smaller in many-body corrections using the PMO semiempirical method, as illustrated for a range of water clusters with and without periodic boundary conditions. In this case, both VMB3 and EMB3 binding energies are nearly in quantitative agreement with the unfragmented QM results (see Supporting Information); however, embedded MB3 is much more efficient than VMB3 in energy calculations since the SCF is carried out on trimer fragments with fixed background charges.

Figure 2 displays the interaction energy due to charge delocalization with respect to charge transferred for the cluster

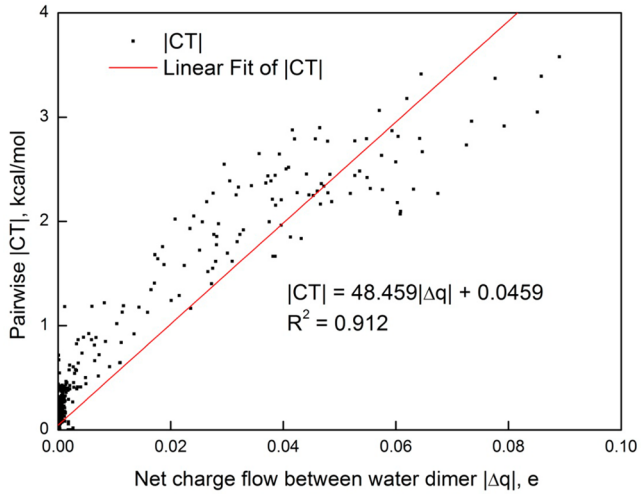


Figure 2. Computed dimeric charge transfer energy versus the amount of charge transferred from a donor water molecule into an acceptor water molecule for a water cluster system consisting of 65 water molecules.

of 65 water molecules at the two-body VMB2 level with HF/6-31G(d). The figure shows a good correlation between charge transfer energy and the net amount of charge migration between two monomers; such correlations are critical to the success of MM force fields, in which charge transfer are not explicitly included. Similar correlations have been observed in cation−π complexes from energy decomposition analysis, and they can be rationalized based on perturbation theory.[48]

3.2. X-Pol as a Quantum Mechanical Force Field

Although *ab initio* molecular orbital theory and density functional theory can be used to improve the accuracy of X-Pol results, it is still impractical to use these methods to perform molecular dynamics simulations on large systems for an extended period of time. With increased computing power, this will become feasible in the future; however, at present, it is desirable to use semiempirical methods such as neglect of diatomic differential overlap (NDDO)[49] methods or the more recent self-consistent-charge density functional tight-binding (SCC-DFTB)[50] method to model condensed-phase systems and biomacromolecules.

The construction of a QMFF based on the variational X-Pol formalism has two components. First, a computationally efficient quantum chemical model is needed to describe the electronic structure of individual molecular fragments. Here we use the PMO method,[51] which is based on the MNDO formalism with the addition of a set of p-orbitals on each hydrogen atom. It was found that the computed molecular polarizabilities for a range of compounds containing hydrogen, carbon, oxygen, and fluorine are significantly improved in this method.[21,51,52] In addition to the enhancement in computed molecular polarizability, a damped dispersion function is included as a post-SCF contribution to the electronic energy.[53] Second, a practical and *parametrizable* procedure is desired to model interfragment electrostatic and exchange-dispersion interactions. For the electrostatic component, a dipole-preserving and polarization consistent (DPPC) method has been used to derive partial atomic charges that can exactly reproduce the instantaneous molecular dipole moment from the polarized electron density of each fragment.[54] Since the DPPC charges are optimized by a Lagrange multiplier technique, there are no new empirical parameters. For the ΔE_{XCD} term, pairwise Lennard-Jones potentials were adopted with monatomic parameters; such potentials contain two parameters for each atomic number.[15,17,20] Employing this strategy, X-Pol quantum chemical models for water[18] and for hydrogen fluoride[21] have been developed for fluid simulations.

3.2.1. Water. The water model is called XP3P, and the system-specific parametrization of the PMO method that is used as part of this model is called PMOw. The results from simulations of liquid water[18] are displayed in Table 4 for 25 °C and 1 atm, along with results from the TIP3P pairwise potential and from two polarizable force fields, namely, AMOEBA and SWM4-NDP. The data in Table 4 were obtained from an assemblage of over 900 million configurations in Monte Carlo simulation and 500 ps of molecular dynamics simulation for a cubic system containing 267 water molecules;[18] the latter took about 24 h (1 fs integration step) on an eight-core workstation, while about 7 million configurations can be executed in Monte Carlo sampling. The average density and heat of vaporization of the XP3P water are within 1% of the experimental values. Quantities involving intermolecular fluctuations such as isothermal compressibility, coefficient of thermal expansion, and dielectric constant are more difficult to converge, but

dx.doi.org/10.1021/ar5002186 | Acc. Chem. Res. 2014, 47, 2837−2845

Table 4. Computed Liquid Properties of the XP3P Model for Water along with Those from Experiments, and the TIP3P, AMOEBA, and SWM4-NDP Models[a]

	XP3P	TIP3P	AMOEBA	SWM4-NDP	expt
ΔH_v, kcal/mol	10.42 ± 0.01	10.41	10.48	10.51	10.51
density, g/cm^3	0.996 ± 0.001	1.002	1.000	1.000	0.997
C_p, cal mol^{-1} K^{-1}	21.8 ± 1.0	20.0	20.9		18.0
$10^6\kappa$, atm^{-1}	25 ± 2	60			46
$10^5\alpha$, K^{-1}	37 ± 3	75			26
μ_{gas}, D	1.88	2.31	1.77	1.85	1.85
μ_{liq}, D	2.524 ± 0.002	2.31	2.78	2.33	2.3−2.6
10^5D, cm^2/s	2.7	5.1	2.02	2.3	2.3
ε	97 ± 8	92	82	79 ± 3	78
τ_D, (ps)	8.8			11 ± 2	8.3
τ_{NMR}, (ps)	2.6			1.87 ± 0.03	2.1

Adapted with permission from ref 18. Copyright 2013 AIP Publishing LLC [a]ΔH_v, heat of vaporization; C_p, heat capacity; κ, isothermal compressibility; α, coefficient of thermal expansion; μ, dipole moment; D, diffusion constant; ε, dielectric constant; τ_D, Debye relaxation time; and τ_{NMR}, NMR rotational relaxation time.

overall, the agreement with experiment is good,[18] and the performance of the XP3P model is as good as any other empirical force field in dynamics simulations.

Monte Carlo simulations of liquid water were performed at temperatures ranging from −40 to 100 °C, and Figure 3 shows

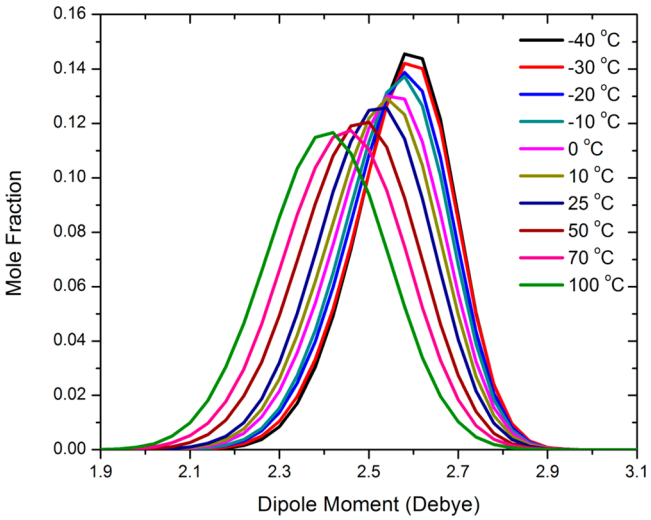


Figure 3. Distribution of the molecular dipole moment of water in the liquid at temperatures ranging from −40 to 100 °C. Adapted with permission from ref 18. Copyright 2013 AIP Publishing LLC.

the distribution of the instantaneous molecular dipole moments. We can infer from the broadness of the observed distribution that water molecules in the liquid experience a wide spectrum of instantaneous electrostatic fields from the rest of the system. Figure 3 indicates that the maximum value in the dipole probability density distribution is the same at different temperatures, despite the large change in the dipole fluctuations, suggesting that there is polarization saturation in liquid water. The average molecular dipole moment of water in the liquid at 25 °C was determined to be 2.524 ± 0.002 D, which represents an increase of 35% relative to the gas-phase value (1.88 D) from the PMOw Hamiltonian. There is no experimental data for direct comparison, but Sprik concluded that an average dipole moment of 2.5−2.6 D in liquid water would most likely yield the correct dielectric constant.[55]

All other thermodynamic and dynamic properties determined using the XP3P model in Table 4 agree reasonably well with experiments and are of similar accuracy to results by other empirical models.[18] Figure 4 shows the structure of liquid water

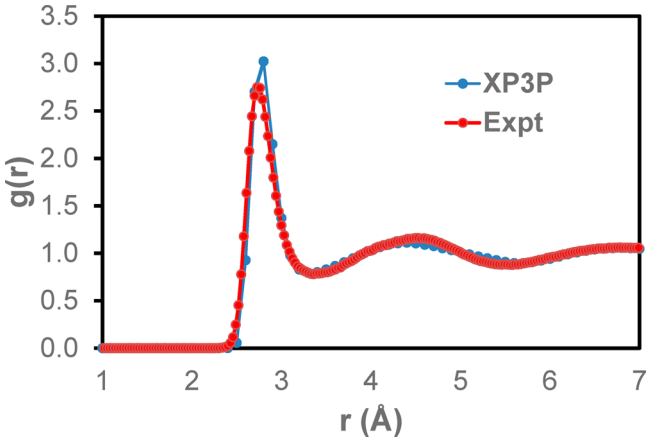


Figure 4. Computed (blue) and experimental (red) radial distribution functions for the O−O pair in liquid water at 25 °C. Adapted with permission from ref 18. Copyright 2013 AIP Publishing LLC.

characterized by radial distribution functions, $g_{xy}(r)$. In comparison with the neutron scattering data,[56] the computational results are in excellent agreement with experiments. In particular, a well-resolved minimum was obtained following the first peak in the O−O distribution, which with the XP3P potential is at 2.78 ± 0.05 Å with a peak height of 3.0. The corresponding experimental values are 2.73 Å and 2.8 from neutron diffraction.[56]

3.2.2. Ionic Liquid. The X-Pol method was used in molecular dynamics simulations for a system consisting of 125 pairs of 1-ethyl-3-methylimidazolium (EMIM) cations and acetate anions in a periodic box using the AM1 model[57] to represent each ion as a molecular fragment. Conventional molecular dynamics simulations of ionic liquids often employ charge-scaled force fields to approximate charge transfer effects between anions and cations, and scaling factors ranging from 0.7 to 0.9 have been used.[58−60] With the VMB2 scheme, it was found that there is a relatively small amount of charge transfer between acetate and EMIM, with an average departure of only 0.03 from unit ionic charges. In contrast, the average charge

transfer over five minimum energy configurations in the gas phase was 0.10, similar to other studies.[58] Thus, it appears that condensed phase polarization effects significantly reduce the amount of charge transfer between cations and anions in ionic liquids; this casts doubt on modeling them with charge-scaling factors significantly deviating from unity.

3.2.3. Protein Simulations. The small protein bovine pancreatic trypsin inhibitor (BPTI) was simulated in water with periodic boundary conditions using the AM1 electronic structure method.[23] The simulation involved 14 281 atoms and 29 026 basis functions, for which 3.2 ps (1 fs time-step) could be performed per day on a single processor (1.66 GHz) computer. Figure 5 depicts the average partial atomic charge of

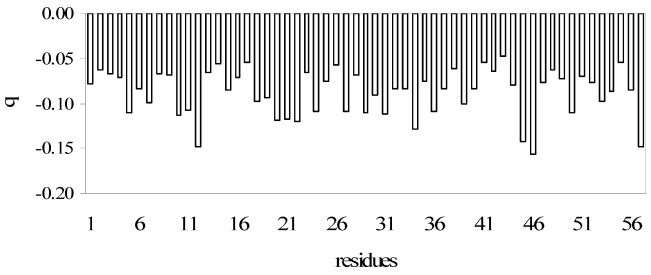


Figure 5. Average partial charges (in atomic units) on the backbone carbonyl (C═O) groups of BPTI, arranged in order of sequence number. Adapted with permission from ref 23. Copyright 2009 American Chemical Society.

the carbonyl group for each amino acid of the polypeptide chain. There is significant fluctuation of the charge distribution in the backbone carbonyl group depending on its sequence location because of its different electrostatic environment as a result of specific hydrogen-bonding interactions with the solvent and polarization by nearby residues. (Since standard semiempirical methods are known to underestimate intermolecular polarization effects, the findings from this study likely represent a lower limit of the polarization effects.) This study illustrates the importance of electronic polarization effects in macromolecular simulations. For comparison, the standard CHARMM22 force field makes use of neutral group convention where the total net atomic charge of the peptide carbonyl group is zero, and the partial atomic charges in the group are the same in all conformational substates.

4. CONCLUDING REMARKS

In this Account, we presented a new strategy to construct potential energy surfaces for macromolecular and liquid-phase systems on the basis of quantum mechanics. Electronic structure theory is directly used to model intermolecular interactions, rather than using quantum mechanical calculations to fit analytical potentials. Potential energy surfaces produced by this approach are called quantum mechanical force fields (QMFFs).

Our strategy, called X-Pol, is based on the partition of condensed-phase and macromolecular systems into fragments to achieve computational efficiency. The intrafragment polarization and the mutual polarization between different fragments are treated explicitly using self-consistent quantum mechanical calculations. The present theory can be used as a single-level method in which all fragments are treated in the same way or more generally as a multilevel electronic structure model in which different fragments are represented by different quantum chemical methods. The X-Pol wave function can be optimized either using a variational Fock operator or through a charge-embedding iterative approach. In both cases, interfragment electrostatic interactions are directly included, whereas the remaining effects, called ΔE_{XCD}, consisting of exchange-repulsion, charge delocalization, and missing dispersion and correlation energies, can be systematically incorporated using a many-body expansion or by approximating ΔE_{XCD} with empirical functions.

So far, we have shown the feasibility of X-Pol for molecular dynamics simulations of a solvated protein in aqueous solution and the accuracy of the X-Pol QMFF for condensed-phase simulations, including liquid water and hydrogen fluoride. The method can also model physical properties that cannot be directly obtained using classical force fields. The project of developing broadly applicable parameters for using X-Pol as a QMFF is still in its infancy, but the method has great promise.

■ ASSOCIATED CONTENT

Supporting Information

A table containing the PMO energies from variational and charge-embedding X-Pol and many-body expansion theory up to three-body terms for water clusters and periodic systems is provided, along with the Cartesian coordinates for a cluster of 65 water molecules. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gao@jialigao.org.

Funding

We thank the National Institutes of Health (Grants GM46376 and GM091445) for partial support of this research.

Notes

The authors declare no competing financial interest.

Biographies

Jiali Gao studied at Beijing University, followed by graduate education at Purdue and postdoctoral research at Harvard. He was on the faculty of the State University of New York at Buffalo, and he is currently a Professor of Chemistry with partial appointments at the University of Minnesota and Jilin University.

Donald G. Truhlar earned a B.A. from St. Mary's College of Minnesota and a Ph.D. from Caltech. He is Regents Professor of Chemistry, Chemical Physics, Nanoparticle Science and Engineering, and Scientific Computation at the University of Minnesota, where he has served on the faculty since 1969.

Yingjie Wang earned a B.S. in chemistry (2009) from Nanjing University. He is currently a fifth-year graduate student of Chemical Physics at University of Minnesota.

Michael J. M. Mazack received a B.S. in mathematics with a physics minor from Western Washington University in 2007. He obtained an M.S. in mathematics from the same institution in 2009, studying numerical linear algebra under Professor Tjalling Ypma. He completed a Ph.D. in scientific computation at the University of Minnesota in 2014 under Professor Jiali Gao, researching the explicit polarization theory and developing coarse-grained models for proteins.

Patrick Löffler earned a B.S. in biology (2009) from the University of Erlangen-Nuremberg and a M.S. in bioinformatics (2012) from the University of Hamburg. After having spent time as a Visiting

Researcher at the University of Minnesota, he is currently a Ph.D. student at the University of Regensburg.

Makenzie Provorse received a B.S. in Chemistry from Kansas State University in 2009 and a M.S. and Ph.D. in Chemistry from the University of Minnesota under the guidance of Prof. Jiali Gao in 2010 and 2014, respectively. She is currently a postdoctoral associate with Assistant Prof. Christine Isborne at the University of California at Merced.

Pavel L. Rehak earned a B.S. from SUNY Stony Brook in 2010, double majoring in Chemistry and Mathematics. He is currently a graduate student under the supervision of Dr. Jiali Gao.

REFERENCES

(1) Levitt, M.; Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **1969**, *46*, 269−279.

(2) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665−6670.

(3) MacKerell, A. D., Jr. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

(4) Aida, M.; Corongiu, G.; Clementi, E. Ab initio force field for simulations of proteins and nucleic acids. *Int. J. Quantum Chem.* **1992**, *42*, 1353−1381.

(5) Gao, J. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* **1997**, *101*, 657−663.

(6) Xie, W.; Gao, J. Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890−1900.

(7) Van der Vaart, A.; Merz, K. M., Jr. The role of polarization and charge transfer in the solvation of biomolecules. *J. Am. Chem. Soc.* **1999**, *121*, 9182−9190.

(8) Giese, T. J.; Chen, H. Y.; Dissanayake, T.; Giambasu, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T. S.; Panteva, M. T.; Radak, B. K.; York, D. M. A variational linear-scaling framework to build practical, efficient next-generation orbital-based quantum force fields. *J. Chem. Theory Comput.* **2013**, *9*, 1417−1427.

(9) Giese, T. J.; Chen, H. Y.; Huang, M.; York, D. M. Parametrization of an orbital-based linear-scaling quantum force field for noncovalent interactions. *J. Chem. Theory Comput.* **2014**, *10*, 1086−1098.

(10) Stoll, H.; Preuss, H. On the direct calculation of localized HF orbitals in molecular clusters, layers and solids. *Theor. Chem. Acc.* **1977**, *46*, 11−21.

(11) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632−672.

(12) Giese, T. J.; Huang, M.; Chen, H.; York, D. M. Recent advances toward a general purpose linear-scaling quantum force field. *Acc. Chem. Res.* **2014**, DOI: 10.1021/ar500103g.

(13) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for benchmark accuracy with the many-body expansion. *Acc. Chem. Res.* **2014**, DOI: 10.1021/ar500119q.

(14) Gao, J.; Wang, Y. Communication: Variational many-body expansion: Accounting for exchange repulsion, charge delocalization, and dispersion in the fragment-based explicit polarization method. *J. Chem. Phys.* **2012**, *136*, No. 071101.

(15) Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys.* **1998**, *109*, 2346−2354.

(16) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field. *J. Chem. Phys.* **2008**, *128*, No. 234108.

(17) Song, L.; Han, J.; Lin, Y. L.; Xie, W.; Gao, J. Explicit polarization (X-Pol) potential using ab initio molecular orbital theory and density functional theory. *J. Phys. Chem. A* **2009**, *113*, 11656−11664.

(18) Han, J.; Mazack, M. J. M.; Zhang, P.; Truhlar, D. G.; Gao, J. Quantum mechanical force field for water with explicit electronic polarization. *J. Chem. Phys.* **2013**, *139*, No. 054503.

(19) Zhang, P.; Truhlar, D. G.; Gao, J. Fragment-based quantum mechanical methods for periodic systems with Ewald summation and mean image charge convention for long-range electrostatic interactions. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7821−7829.

(20) Han, J. B.; Truhlar, D. G.; Gao, J. L. Optimization of the explicit polarization (X-Pol) potential using a hybrid density functional. *Theor. Chem. Acc.* **2012**, *131*, 1161−1167.

(21) Mazack, M. J. M.; Gao, J. Quantum mechanical force field for hydrogen fluoride with explicit electronic polarization. *J. Chem. Phys.* **2014**, *140*, No. 204501.

(22) IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and symbols for description of conformation of polypeptide. *Pure Appl. Chem.* **1974**, *40*, 291−308.

(23) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. X-Pol potential: An electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *J. Chem. Theory Comput.* **2009**, *5*, 459−467.

(24) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* **1998**, *102*, 4714−4721.

(25) Gao, J. Hybrid quantum mechanical/molecular mechanical simulations: An alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* **1996**, *29*, 298−305.

(26) Singh, U. C.; Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH3Cl + Cl- exchange reaction and gas phase protonation of polyenes. *J. Comput. Chem.* **1986**, *7*, 718−730.

(27) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. Incorporation of QM/MM buffer zone in the variational double self-consistent field method. *J. Phys. Chem. B* **2008**, *112*, 14124−14131.

(28) Leverentz, H. R.; Gao, J. L.; Truhlar, D. G. Using multipole point charge distributions to provide the electrostatic potential in the variational explicit polarization (X-Pol) potential. *Theor. Chem. Acc.* **2011**, *129*, 3−13.

(29) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic study of the embedding potential description in the fragment molecular orbital method. *J. Phys. Chem. A* **2010**, *114*, 8742−8753.

(30) Wang, B.; Yang, K. R.; Xu, X.; Isegawa, M.; Leverentz, H. R.; Truhlar, D. G. Quantum mechanical fragment methods based on partitioning atoms or partitioning coordinates. *Acc. Chem. Res.* **2014**, DOI: 10.1021/ar500068a.

(31) Saha, A.; Raghavachari, K. Dimers of dimers (DOD): A new fragment-based method applied to large water clusters. *J. Chem. Theory Comput.* **2014**, *10*, 58−67.

(32) Senn, H. M.; Thiel, W. QM/MM methods for biological systems. *Top. Curr. Chem.* **2007**, *268*, 173−290.

(33) Wang, Y. J.; Sosa, C. P.; Cembran, A.; Truhlar, D. G.; Gao, J. L. Multilevel X-Pol: A fragment-based method with mixed quantum mechanical representations of different fragments. *J. Phys. Chem. B* **2012**, *116*, 6781−6788.

(34) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.

(35) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully analytic energy gradient in the fragment molecular orbital method. *J. Chem. Phys.* **2011**, *134*, No. 124115.

(36) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. The fragment molecular orbital method for geometry optimizations of polypeptides and proteins. *J. Phys. Chem. A* **2007**, *111*, 2722−2732.

(37) Richard, R. M.; Lao, K. U.; Herbert, J. M. Approaching the complete-basis limit with a truncated many-body expansion. *J. Chem. Phys.* **2013**, *139*, No. 224102.

(38) Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar D, G.; Gao, J. On the interfragment exchange in the X-Pol method. *J. Chem. Theory Comput.* **2010**, *6*, 2469−2476.

(39) Gao, J.; Cembran, A.; Mo, Y. Generalized X-Pol theory and charge delocalization states. *J. Chem. Theory Comput.* **2010**, *6*, 2402−2410.
(40) Mo, Y. R.; Bao, P.; Gao, J. L. Energy decomposition analysis based on a block-localized wavefunction and multistate density functional theory. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6760−6775.
(41) Lao, K. U.; Herbert, J. M. An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method. *J. Chem. Phys.* **2013**, *139*, No. 034107.
(42) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New advance in computational chemistry: Full quantum mechanical ab initio computation of streptavidin−biotin interaction energy. *J. Phys. Chem. B* **2003**, *107*, 12039−12041.
(43) Li, W.; Li, S.; Jiang, Y. Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.
(44) Truhlar, D. G.; Dahlke, E. E.; Leverentz, H. R. Evaluation of the electrostatically embedded many-body expansion and the electrostatically embedded many-body expansion of the correlation energy by application to low-lying water hexamers. *J. Chem.Theory Comput.* **2008**, *4*, 33−41.
(45) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. QM:QM electronic embedding using Mulliken atomic charges: Energies and analytic gradients in an ONIOM framework. *J. Chem. Phys.* **2008**, *128*, No. 034107.
(46) Fedorov, D. G.; Kitaura, K. Second order Møller-Plesset perturbation theory based upon the fragment molecular orbital method. *J. Chem. Phys.* **2004**, *121*, 2483−2490.
(47) Wang, B.; Truhlar, D. G. Including charge penetration effects in molecular modeling. *J. Chem. Theory Comput.* **2010**, *6*, 3330−3342.
(48) Mo, Y.; Subramanian, G.; Gao, J.; Ferguson, D. M. Cation-π interactions: An energy decomposition analysis and its implication in δ-opioid receptor-ligand binding. *J. Am. Chem. Soc.* **2002**, *124*, 4832−4837.
(49) Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate self-consistent molecular orbital theory. I. Invariant procedures. *J. Chem. Phys.* **1965**, *43*, S129−S135.
(50) Elstner, M.; Porezag, D.; Juugnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Sukai, S.; Seifect, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(51) Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. L. Polarized molecular orbital model chemistry. 2. The PMO method. *J. Chem. Theory Comput.* **2011**, *7*, 857−867; **2012**, *8*, 2983.
(52) Isegawa, M.; Fiedler, L.; Leverentz, H. R.; Wang, Y. J.; Nachimuthu, S.; Gao, J. L.; Truhlar, D. G. Polarized Molecular Orbital Model Chemistry 3. The PMO Method Extended to Organic Chemistry. *J. Chem. Theory Comput.* **2013**, *9*, 33−45.
(53) McNamara, J. P.; Sharma, R.; Vincent, M. A.; Hillier, I. H.; Morgado, C. A. The non-covalent functionalisation of carbon nanotubes studied by density functional and semi-empirical molecular orbital methods including dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 128−135.
(54) Zhang, P.; Bao, P.; Gao, J. L. Dipole preserving and polarization consistent charges. *J. Comput. Chem.* **2011**, *32*, 2127−2139.
(55) Sprik, M. Hydrogen bonding and the static dielectric constant in liquid water. *J. Chem. Phys.* **1991**, *95*, 6762−6769.
(56) Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **2000**, *258*, 121−137.
(57) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(58) Morrow, T. I.; Maginn, E. J. Molecular dynamics study of the ionic liquid 1-n-butyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Chem. B* **2002**, *106*, 12807−12813.
(59) Youngs, T. G. A.; Hardacre, C. Application of static charge transfer within an ionic-liquid force field and its effect on structure and dynamics. *ChemPhysChem* **2008**, *9*, 1548−1558.
(60) Wendler, K.; Zahn, S.; Dommert, F.; Berger, R.; Holm, C.; Kirchner, B.; Delle Site, L. Locality and fluctuations: Trends in imidazolium-based ionic liquids and beyond. *J. Chem. Theory Comput.* **2011**, *7*, 3040−3044.