Recent Advances toward a General Purpose Linear-Scaling Quantum Force Field

Timothy J. Giese,[†] Ming Huang,[†,‡] Haoyuan Chen,[†] and Darrin M. York*,[†]

[†]Center for Integrative Proteomics Research, BioMaPS Institute, and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854-8087 United States

[‡]Scientific Computation, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455−0431, United States

CONSPECTUS: There is need in the molecular simulation community to develop new quantum mechanical (QM) methods that can be routinely applied to the simulation of large molecular systems in complex, heterogeneous condensed phase environments. Although conventional methods, such as the hybrid quantum mechanical/molecular mechanical (QM/MM) method, are adequate for many problems, there remain other applications that demand a fully quantum mechanical approach. QM methods are generally required in applications that involve changes in electronic structure, such as when chemical bond formation or cleavage occurs, when molecules respond to one another through polarization or charge

transfer, or when matter interacts with electromagnetic fields. A full QM treatment, rather than QM/MM, is necessary when these features present themselves over a wide spatial range that, in some cases, may span the entire system. Specific examples include the study of catalytic events that involve delocalized changes in chemical bonds, charge transfer, or extensive polarization of the macromolecular environment; drug discovery applications, where the wide range of nonstandard residues and protonation states are challenging to model with purely empirical MM force fields; and the interpretation of spectroscopic observables. Unfortunately, the enormous computational cost of conventional QM methods limit their practical application to small systems. Linear-scaling electronic structure methods (LSQMs) make possible the calculation of large systems but are still too computationally intensive to be applied with the degree of configurational sampling often required to make meaningful comparison with experiment.

In this work, we present advances in the development of a quantum mechanical force field (QMFF) suitable for application to biological macromolecules and condensed phase simulations. QMFFs leverage the benefits provided by the LSQM and QM/MM approaches to produce a fully QM method that is able to simultaneously achieve very high accuracy and efficiency. The efficiency of the QMFF is made possible by partitioning the system into fragments and self-consistently solving for the fragment-localized molecular orbitals in the presence of the other fragment's electron densities. Unlike a LSQM, the QMFF introduces empirical parameters that are tuned to obtain very accurate intermolecular forces. The speed and accuracy of our QMFF is demonstrated through a series of examples ranging from small molecule clusters to condensed phase simulation, and applications to drug docking and protein−protein interactions. In these examples, comparisons are made to conventional molecular mechanical models, semiempirical methods, *ab initio* Hamiltonians, and a hybrid QM/MM method. The comparisons demonstrate the superior accuracy of our QMFF relative to the other models; nonetheless, we stress that the overarching role of QMFFs is not to supplant these established computational methods for problems where their use is appropriate. The role of QMFFs within the toolbox of multiscale modeling methods is to extend the range of applications to include problems that demand a fully quantum mechanical treatment of a large system with extensive configurational sampling.
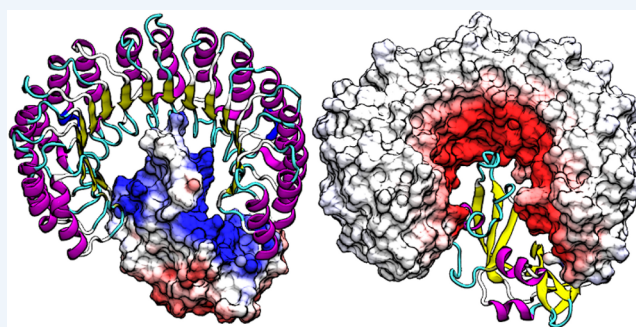
1. INTRODUCTION

Computational modeling plays a vital role in chemical research by aiding in the interpretation of experimental measurements, guiding the design of future experiments, and making predictions when experiment is unavailable. The variety and complexity of many chemical problems, such as those encountered in biology, require an array of computational tools ranging from molecular mechanical (MM) force fields to quantum mechanical (QM) methods to probe the dynamical, reactive, and electromagnetic phenomenon of interest.

Theoretical methods rely upon inherent approximations that limit their accuracy, range of application, and computational efficiency. Computational modeling therefore begins with the selection of the most appropriate tool for the task at hand. Conventional MM force fields are useful for problems that
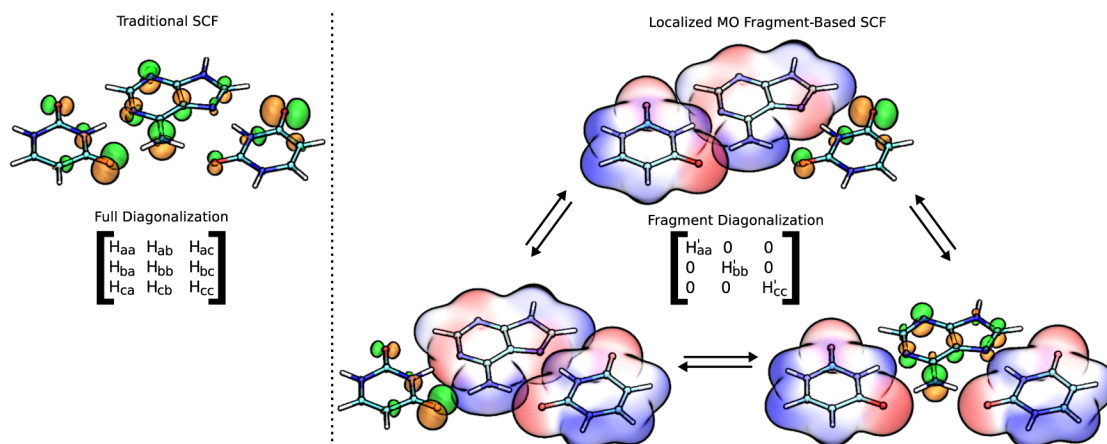
Figure 1. Illustration of the difference between a standard QM calculation and a MO-based QMFF. The traditional SCF constructs MOs spanning the space of the entire system by diagonalizing a large Fock matrix. The QMFF MOs are localized on each fragment and are obtained from the diagonalization of a series of small Fock matrices. The blue and red molecular surfaces are meant to represent the external chemical potential experienced by the active fragment. The arrows represent the equilibration of the fragment systems due their coupling through their electron densities.

require a large amount of configurational sampling, but they are not designed to describe bond formation and cleavage. On the other hand, QM methods can accurately model chemical reactions (depending on the choice of Hamiltonian and basis set), but they are too costly for generating ensemble statistics. Hybrid QM/MM methods treat a small part of the system with a QM model and the remainder with a MM force field. QM/MM thus offers a practical compromise to enable a direct comparison with experiment through simulation of chemical reactions;[1,2] however, the QM/MM approach inherits all the problems inherent within the QM and MM models and introduces new challenges involving their interaction. For example, standard MM models neglect multipolar electrostatics and many-body polarization; the balance between affordability and accuracy of the QM method is still an issue; the accuracy of the QM nonbonded interactions is questionable, especially when large QM regions are required; and the QM/MM interaction potential does not properly adjust as the reaction proceeds, as would be expected when significant changes in atomic charges occur.[3]

Quantum mechanical force fields (QMFFs) are a class of methods that divide a system into fragments, each of which are treated quantum mechanically but whose interactions are empirically modeled to recover high accuracy while remaining computationally efficient. QMFFs are not meant to replace well-established methods such as QM/MM for problems where those methods are appropriate; rather, their purpose is to extend the range of applications to include problems that demand a full QM description of a large system requiring extensive configurational sampling. Example application areas that stand to benefit from QMFFs include (1) enzyme design studies of catalytic events that involve delocalized changes in chemical bonds, charge transfer, or extensive polarization of the macromolecular environment, thus requiring an extended QM region; (2) drug discovery applications, where the wide range of nonstandard residues and protonation states are challenging to model with purely empirical MM force fields; and (3) interpretation of spectroscopic observables of biological macromolecules, for example, infrared (IR), Raman,[4] nuclear magnetic resonance[5−7] (NMR), and 2-D IR[8] spectra, which are inherently quantum mechanical in nature.

A variety of QMFFs have been examined through proof-of-principle studies that introduce the methodology and demonstrate their feasibility.[9−14] The ultimate success of a model, however, is judged through its application and subsequent assessment of its balance between *speed* and *accuracy* relative to established techniques.[15] The first stages of model development therefore involve a substantial level of effort to produce results used to assess the advantages and disadvantages of proposed models. To this end, the present work highlights recent advances that demonstrate the speed and accuracy of a QMFF relative to standard approaches. The relationship between QMFFs and traditional QM approaches (including linear-scaling electronic structure methods) is discussed, and we explain how those differences are exploited by QMFFs to achieve their tremendous computational speed-up while simultaneously delivering superior accuracy.

2. BACKGROUND

Linear-scaling quantum methods (LSQMs), that is, methods whose computational complexity increases linearly with system size, surmount the high cost of traditional QM algorithms by avoiding the construction of globally orthonormal molecular orbitals (MOs). Fragment-based LSQMs, for example, achieve linear scaling by subdividing the system into fixed-size regions of locally orthogonal orbitals[16] that are constructed from the diagonalization of matrices whose sizes are proportional to their respective fragments, so although the number of fragments increase with system size, the complexity required to generate each set of fragment MOs remains constant. The various fragment-based LSQMs, such as the divide-and-conquer[17] and fragment molecular orbital[18] methods, represent the different approaches used to correct the short-ranged interfragment interactions to account for having relaxed the global orthogonality constraints. Although there are technical differences between the methods, their corrections act to mimic the effect of having enforced the MO orthogonality of a fragment with its neighbors, which we refer to as the fragment's "buffer". By extending the size of the buffer, LSQMs are capable of well-reproducing the result of a traditional implementation without introducing new, empirically parametrized corrections; however, the use of a buffer necessarily cause LSQMs to become

computationally advantageous only when applied to large systems. One will, therefore, pragmatically choose to use them with relatively inexpensive QM Hamiltonians, for example, semiempirical models or density functional theory (DFT) methods with small basis sets. In this event, the capability of reproducing the result of a standard implementation is not necessarily advantageous because the cost savings provided by inexpensive Hamiltonians are countered by their poor representation of nonbonded interactions.

Orbital-based QMFFs are designed to circumvent the limitations of LSQM methods by replacing some of their theoretical rigor with computationally tractable models that can be tuned emprically to achieve high accuracy. First, the "break-even point" at which the method becomes computationally advantageous to use is eliminated by removing the fragment buffers entirely. Second, parametrized nonbond interactions are introduced to account for the lack of explicit interfragment orbital coupling. Furthermore, the empirical parametrization affords the opportunity to improve the description of nonbond interactions beyond the capabilities of the underlying QM model. As illustrated in Figure 1, the QMFF calculation of the fragment MOs is *nearly* equivalent to having solved a series of small, independent *ab initio* calculations upon embedding the fragment within the remainder of the system, which is viewed, from the fragment's perspective, merely as a source of an external potential. Although there are no explicit interfragment MO coupling matrix elements, the MOs of each fragment are coupled with the others through the interaction of their electron densities, which change at each self-consistent field (SCF) step until a global convergence is reached.

There exists an alternative, promising strategy for constructing QMFFs that avoids the use of MOs entirely. These "density-based QMFFs" prefit an accurate *ab initio* electron density with an auxiliary basis that interacts through a density-overlap model and is allowed to respond under the principle of chemical potential equalization.[13,14] This class of QMFFs, however, is not designed to model changes in chemical bonding and is not the focus of this Account.

3. DEVELOPMENT OF A QMFF BASED ON THE mDC METHOD

The orbital-based QMFF strategy described above was pioneered by Gao with the X-Pol model[9,10,19−22] and was adopted in our own method, called mDC.[12,23] mDC and X-Pol are conceptually equivalent but differ in their details of how the fragments interact with one another. The X-Pol method interacts the fragments using a traditional QM/MM potential; however, this potential depends upon which fragment is considered to be the QM region. X-Pol therefore computes the QM/MM potential for each fragment and averages the interaction energy and forces. The X-Pol strategy used to compute interfragment interactions may be sufficient for some Hamiltonians, like neglect of diatomic differential overlap (NDDO)-based semiempirical models, but we have found it to be wanting when applied to popular density functional tight-binding (DFTB) models, like the DFTB3 semiempirical model.[24] A QM/MM treatment of DFTB3 would result in the interaction of atomic charges only; however, DFTB3's ability to well-reproduce hydrogen bond geometries largely results from its explicit coupling of MOs between the molecules, not from the interaction of their atomic charges.

The explicit MO couplings are removed in the QMFF, so we've found it necessary to construct an auxiliary set of atomic

multipole moments from the DFTB3 density matrix to preserve the angular dependence that would normally be expressed from the MOs, as illustrated in Figure 2. Considering that we are
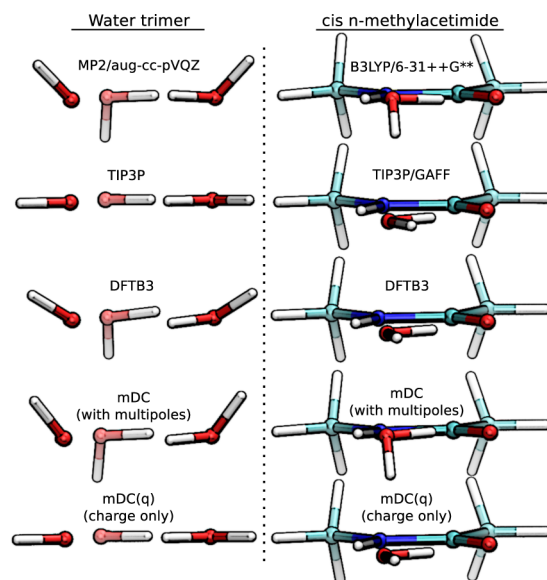

Figure 2. Comparison of hydrogen bond angles produced by different methods. (left) In-plane view of the "up−up−down" water trimer. (right) Hydrogen bond between water and the secondary amine of *n*-methylactimide.

now concocting a new interfragment potential, we choose it to yield symmetric interactions so that, unlike the X-Pol potential, they are computed once, thereby avoiding a need for averaging. By having chosen a symmetric interaction potential, the mDC total energy becomes

$$E_{mDC} = \sum_{A}^{N_{frag}} E_A(\mathbf{R}_A, N_A, \mathbf{p}) + E_{inter}(\mathbf{R}, \mathbf{q}) \tag{1}$$

where E_A is the electronic energy of fragment A with N_A electrons and atom positions $\mathbf{R}_A$ under the influence of the external potential $\mathbf{p} = \partial E_{inter}(\mathbf{R}, \mathbf{q})/\partial \mathbf{q}$, and E_{inter} is the total interfragment interaction energy computed from the atomic multipole moments $\mathbf{q}$.

In choosing a QM model around which a QMFF is to be parametrized, one must consider not only its computational efficiency but also its ability to strike a proper balance between the quality of intra- and interfragment interactions. Semiempirical Hamiltonians are, at the present time, the only models we consider to be sufficiently fast to make the routine application of a QMFF to the dynamics of large systems practical. These methods are faster than *ab initio* Hamiltonians because they use very small AO basis sets, they ignore many of the AO integrals, and they parametrize the remaining integrals to achieve reasonably good geometries, bond enthalpies, and other small molecule properties. Both NDDO and DFTB are established semiempirical methodologies that have been successfully used in the past, but there are some important differences between them. NDDO methods use atom-centered multipoles to perform electrostatics, whereas DFTB3 limits the second-order electrostatics to atomic charges only. On the other hand, NDDO methods ignore interatomic AO overlap, whereas DFTB3 does not. DFTB3's explicit treatment of AO overlap affords a better description of intramolecular

interactions such as ring puckers,[25] as illustrated in Figure 3. We have therefore chosen to use DFTB3 as the base QM
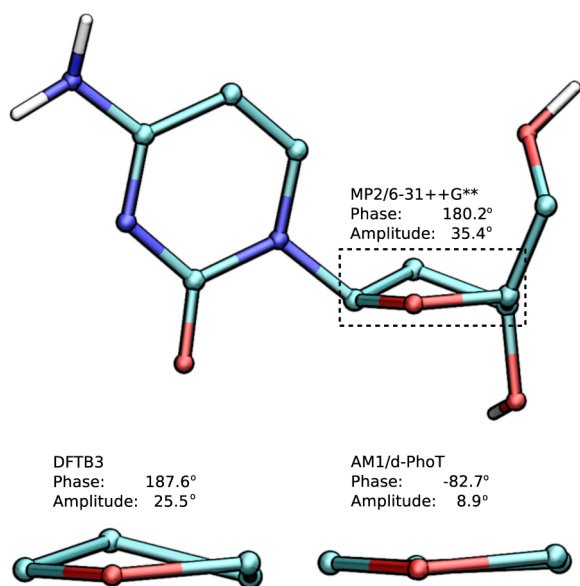

Figure 3. Deoxycytidine. Pucker phase and amplitudes are listed in degrees. Standard torsion constraints are employed to mimic the nucleoside connection to the B-DNA backbone.

model to leverage its description of intramolecular interactions and then parametrically construct atomic multipoles from the DFTB3 density matrix to interact the fragments. By limiting the use of the atomic multipoles to the interfragment interactions only, we can target the parametrization toward nonbond forces without directly affecting the properties of isolated fragments, so we do not have to redevelop a new DFTB3 model from scratch. With this approach, we have recently parametrized an

mDC model that improves electrostatic potentials (Figure 4) and interaction energies.[23] Without having included second-order multipolar electrostatics, DFTB3 shows discrepancies in its electrostatic potentials for sp^3 oxygens, sp^2 nitrogens, and π-bond electrons. Preliminary tests with a point-charge DFTB3 QMFF, called mDC(q) (ref 12), suggests that higher-order multipoles are necessary to retain good hydrogen bond angles (Figure 2), which otherwise devolve into MM force field-like configurations. The electrostatic potentials and hydrogen bond angles are improved upon including higher-order multipoles.

4. ACCURACY OF INTERMOLECULAR INTERACTIONS

In ref 12, we introduced a general linear-scaling QMFF framework that highlighted many of the ingredients we felt might eventually be brought to bear in the development of a full-fledged QMFF. One of those ingredients included a charge-dependent density overlap van der Waals model,[3] similar in spirit to those used in the density-based QMFFs.[13] In most chemical environments, however, significant change in atomic charges will primarily be limited to those atoms involved in chemical reactions. A simple Lennard-Jones (LJ) model should therefore suffice for the majority of a system, and this was the strategy taken in ref 23. In that work, the LJ interactions were parametrized to a number of nonbond databases constructed from benchmark *ab initio* calculations and comparisons were made to a standard MM model (GAFF), NDDO and DFTB semiempirical models, B3LYP/6-31G*, and the MNDO/GAFF hybrid QM/MM method. A small representative subset of those comparisons are shown in Figure 5, which we have supplemented with M062X/6-311++G** results to make comparison with a modern density functional method and large basis set. The comparisons reveal that the GAFF force field produces very accurate nonbond interactions when applied to small systems. B3LYP/6-31G* and MNDO/GAFF are of comparable quality and are both outperformed by GAFF.
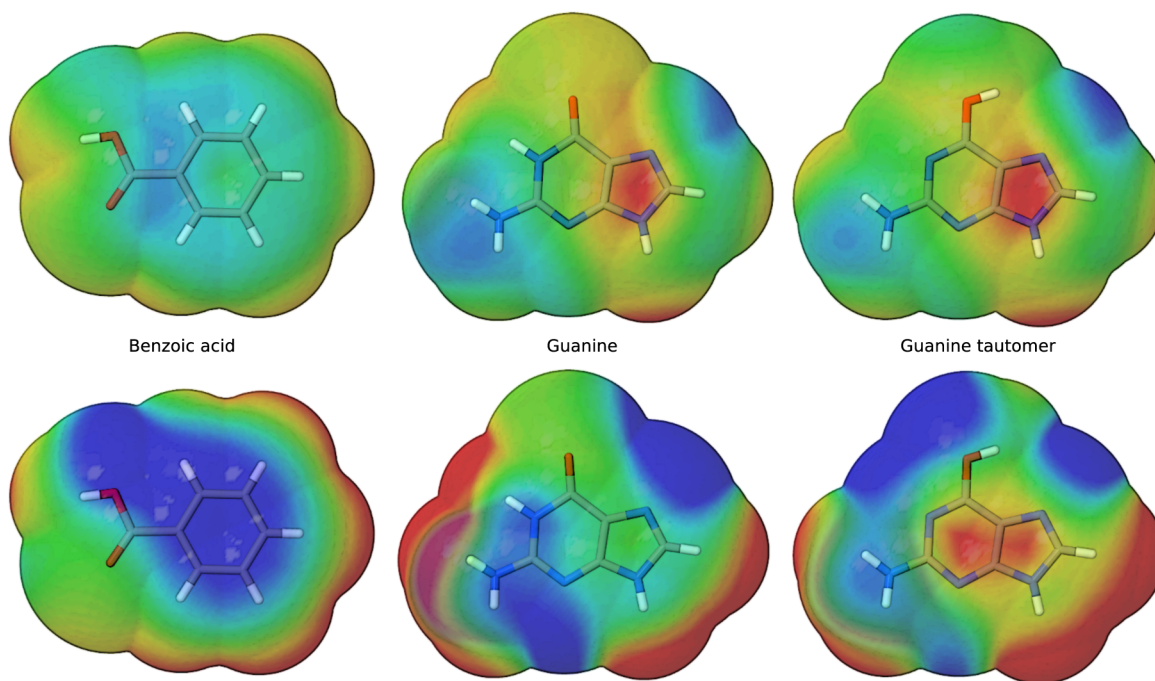

Figure 4. Difference between mDC (top) and DFTB3 (bottom) electrostatic potentials relative to B3LYP/6-311++G**. Colors bounded by ±0.003 au. Blue, red, and green indicate electron density deficiency, excess, and agreement relative to B3LYP/6-311++G**, respectively.
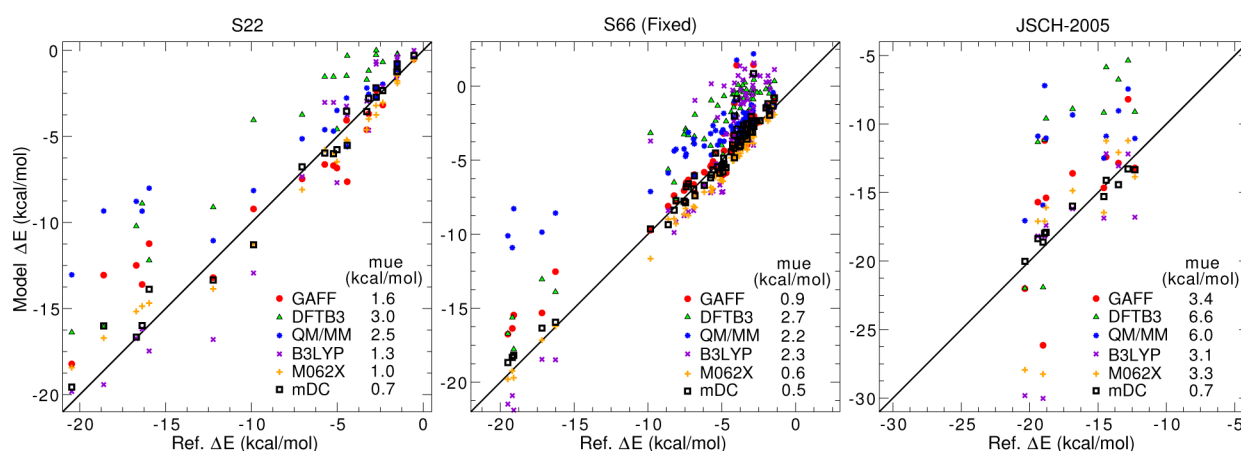
Figure 5. Comparison of mDC with other Hamiltonians using several databases (S22 and JSCH-2005, ref 37; S66, ref 38) of intermolecular interactions. B3LYP, M062X, and QM/MM refer to B3LYP/6-31G*, M062X/6-311++G**, and MNDO/GAFF, respectively. "(Fixed)" indicates that the dimer interaction energies are computed using the reference structures.
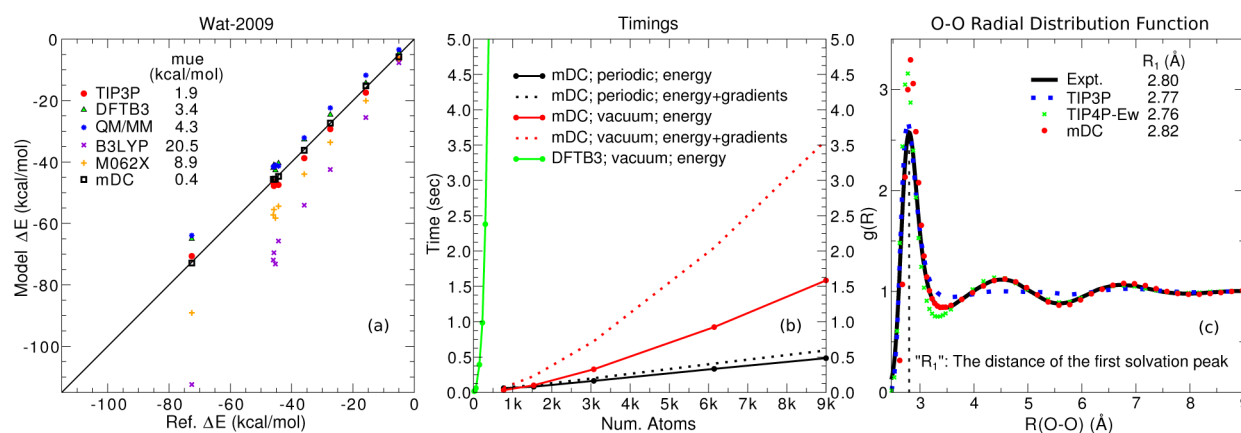

Figure 6. (a) Water cluster relative energies (reference data taken from ref 39). QM/MM refers to MNDO/TIP3P. (b) mDC and DFTB3 water box timings with and without gradient evaluation in gas phase and under periodic boundary conditions. (c) Comparison of condensed phase water O−O radial distribution functions (experimental data taken from ref 40).

DFTB and NDDO semiempirical models were also observed to be of comparable accuracy to each other[23] (the unsigned errors of both are approximately twice those of GAFF), in agreement with previous findings.[26] The parametrized mDC model, however, is found to produce interaction energy mean unsigned errors smaller than any of the MM, QM, or semiempirical QM/MM methods.

5. APPLICATION TO DRUG SCREENING

GAFF's accuracy is impressive when applied to small molecules; however, it is a fixed-charge, nonpolarizable model, so one may therefore question to what extent its quality may degrade when applied to larger systems. Previous studies have concluded, for example, that electronic polarization can stabilize transition state barriers by 9 kcal/mol[27] and that it plays a key role in the relative ordering of inhibitor interactions with focal adhesion kinase[28] and the binding of ligands to the trypsin protein.[29] We recently applied GAFF, MNDO/MM, and mDC to a drug screening exercise involving cyclin-dependent kinase 2 (CDK2).[30] In this simple exercise, drug ligands with known experimental protein inhibition constants (IC_{50}) were docked into the receptor pocket of CDK2, and the gas phase protein−ligand interaction energy was correlated to the experimental IC_{50}'s. We found that the GAFF interaction

energies correlate less to experiment than either MNDO/MM or mDC, and we further found mDC to correlate best. Given that MNDO/MM and mDC both explicitly treat electronic polarizability and both correlate better than the nonpolarizable GAFF, it is reasonable to suspect that this difference is due to explicit polarization. Although reasonable correlations between binding enthalpy and binding free energy are attainable for this system without having to explicitly consider the entropic contributions of the kinase, ligand, or solvent degrees of freedom through thermodynamic averaging via molecular dynamics simulation, it has been previously noted that the correlation between changes in binding enthalpy and free energy can often be poor.[31] The efficiency of QMFFs make feasible the prospect of modeling free energy changes, including entropic effects, with molecular dynamics simulation.

6. CONDENSED PHASE WATER SIMULATIONS

New models seek to attain maximum transferability, that is, the model should produce acceptable results in a variety of environments, so as to widen their range of application. One measure of transferability is the reproduction of both small gas-phase cluster interactions and condensed-phase properties. Figure 6a,c demonstrates the ability of mDC to reproduce both small-to-medium sized water cluster binding energies and the
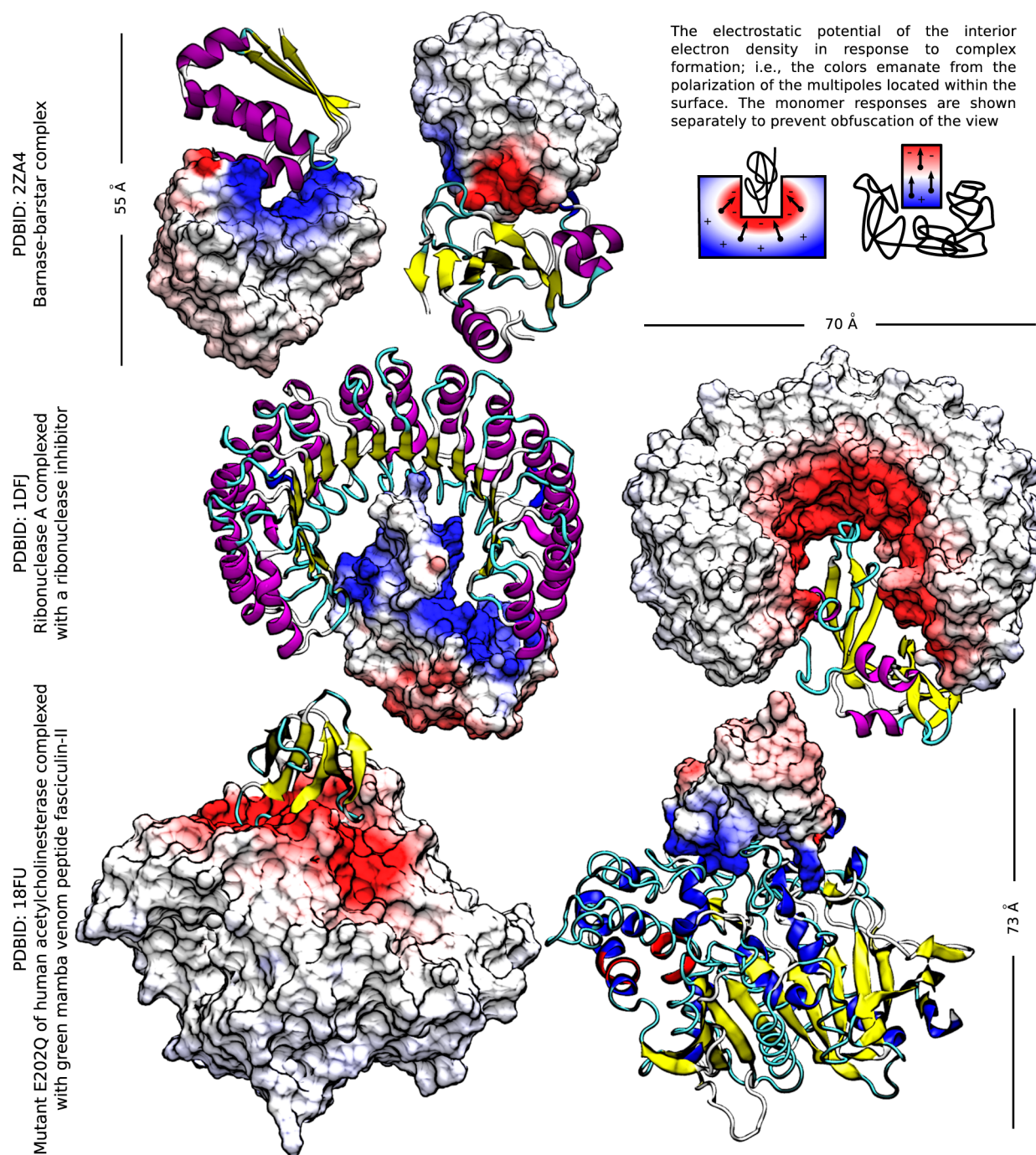
Figure 7. Electrostatic potential arising from the polarization density induced upon protein−protein binding in the gas phase. Blue and red indicate the positive and negative electrostatic potential (bounded by ±0.04 au) from the response of the multipoles interior to the surface. The surface of each monomer is shown independently to aid the visualization.

O−O radial distribution function (RDF) of liquid water obtained from molecular dynamics simulation. The mDC ΔE mean unsigned error is considerably better than the MM TIP3P model, the DFTB3 semiempirical model, the MNDO/TIP3P QM/MM method, and the two *ab initio* Hamiltonians. Both mDC and TIP4P-Ew are found to well-reproduce the experimental RDF. The simulations were performed with a modified version of the SANDER program and consisted of 512 waters sampled for 3.2 ns in a cubic box with a 1 fs time-step in the canonical (NVT) ensemble at a temperature of 298 K and density of 0.996 g/cm³.

Figure 6b compares execution times between mDC in vacuum and periodic boundary conditions and a standard SCF implementation of DFTB3. The standard SCF implementation of DFTB3 would take approximately 24 h to evaluate the energy of a 3000 water system, whereas a condensed phase mDC calculation takes 0.5 s. The 3.2 ns mDC simulation used to construct the RDF required 37 h/ns of simulation. Previous timing analysis of large systems concluded that a point-charge-only variant of mDC is twice as fast.[12]

7. POLARIZATION EFFECTS IN MACROMOLECULAR BINDING

An important gauge of a model's transferability is its ability to accurately describe molecular interactions over a wide range of heterogeneous electrostatic environments.[32] This is particularly important in macromolecular association where electrostatic complementarity is critical for molecular recognition and is often a driving force for binding. QMFFs have the advantage of including a multipolar description of electrostatic interactions as well as explicit many-body polarization. Figure 7 illustrates the importance of these effects by showing the electrostatic potential due to the electronic *polarization density* that occurs upon protein–protein binding for several systems described in more detail below.

The barnase–barstar complex[33] is a classic example of electrostatic complementarity at the binding interface. It is clear that the polarization induced upon complexation enhances both electrostatic complementarity and binding.

Figure 7 displays ribonuclease A (RNase A) bound within the concave cleft of ribonuclease inhibitor (RI),[34] a highly flexible protein consisting of leucine-rich repeats. The cleft consists of an inner layer of parallel β sheets and loops surrounded by an outer layer of parallel α helices. The intermolecular interaction is largely electrostatic, and the positively charged RNase A induces considerable electronegative potential on the concave surface of the RI to enhance binding.

The binding of the E202Q mutant of human acetylcholinesterase (AChE) complexed with green mamba venom peptide fasciculin-II (FAS-II)[35] is shown in the bottom of Figure 7. AChE is a highly efficient enzyme that hydrolyzes the neurotransmitter acetylcholine to terminate synaptic transmission, is highly sensitive to reactions with organophosphorus inhibitors, including nerve agents such as sarin, and is the target of neurotoxins contained in certain snake venoms. Despite its relative small size, FAS-II induces a large electronegative polarization density and potential at the surface of AChE that helps it to achieve its binding strength.

In the above examples, it is clear that the polarization induced by protein–protein binding is significant; it enhances the electrostatic complementarity and thus increases the binding energies. Nonetheless, fixed-charge force fields may overestimate the binding energies due to "pre-polarization" of their charges to account for their lack of an explicit treatment for polarization.[29]

8. SUMMARY AND OUTLOOK

This Account has presented advances in the development of a QMFF to be used as a tool in the study of complex biochemical problems. The efficiency of the QMFF is achieved by limiting the explicit use of MOs to the intrafragment interactions while modeling the interfragment interactions through MO-derived atomic densities and empirically parametrized functions. This decomposition and subsequent parametrization of the interactions allows one to achieve very accurate intermolecular forces. The mDC QMFF is based on the recently developed DFTB3 method, which models relative conformational energies and barriers more reliably than the conventional NDDO semiempirical methods lacking explicit "orthogonalization corrections".

The mDC method presented here is demonstrated to be extremely fast and, in several applications, is found to deliver higher accuracy than conventional molecular mechanical force fields, semiempirical QM/MM potentials, and full QM results. mDC was further shown to improve the representation of molecular electrostatic potentials relative to DFTB3, and its transferability was demonstrated by its ability to reproduce small-to-medium sized water cluster binding energies and the O–O RDF of liquid water. The CDK2 drug screening exercise was used to demonstrate mDC's transferability to heterogeneous environments; mDC was found to better correlate the receptor–ligand binding energies to experimental IC$_{50}$'s than either GAFF or MNDO/MM. The mDC method was applied to protein–protein interactions to illustrate the importance of explicit polarization in protein binding. Finally, the mDC method was shown to be very efficient. On an 8-core desktop workstation, the calculation of the mDC energy and forces of a 9000 atom system requires 0.5 s.

Our results are meant to highlight mDC's accuracy and efficiency to emphasize its promise as a useful tool. These demonstrations, however, have focused solely on nonbonded intermolecular interactions and did not illustrate its ability to be used in chemical reactions or for the calculation of inherently quantum mechanical observables, such as molecular spectra. It thus remains to broadly test and apply the mDC method to study chemical reactions in complex homogeneous and heterogeneous condensed phase environments, such as those encountered in biocatalysis, and to use it as an aid in the interpretation and prediction of 1D- and 2D-IR, Raman, and NMR spectra of biomolecules.

The mDC method provides a general framework from which new QMFFs can be built using higher-level QM methods. Because the coupling between fragments occurs through the electron density, multilevel QMFFs can be designed that combine different theoretical models to construct a global potential that is tailored to optimally balance accuracy and efficiency for a particular problem,[36] and in this way, multilevel QMFFs will provide a set of tools that allow a hierarchy of accuracy and robustness that can be used together in a multiscale modeling framework to solve complex biochemical problems.

The mDC method provides the foundation from which free energy surfaces can be systematically corrected to higher levels. The impact of new QMFFs will be made more powerful still as other technologies evolve, for example, "free energy surface (FES) correction" methods. FES correction methods provide a mechanism whereby complex FESs constructed by exhaustive sampling with an affordable Hamiltonian can be systematically and variationally corrected to closely approximate higher-level surfaces with orders of magnitude reduction in the required sampling. In conclusion, QMFFs provide the foundation for the design of multiscale modeling strategies that, together with new FES methods, will evolve into powerful tools that can be brought to bear on a wide range of biological problems.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: york@biomaps.rutgers.edu.

Notes

The authors declare no competing financial interest.

Biographies

Timothy J. Giese is an Associate Research Professor at the BioMaPS Institute for Quantitative Biology at Rutgers.

Ming Huang is a graduate student in the field of Scientific Computation at the University of Minnesota.

Haoyuan Chen is a chemistry graduate student at Rutgers.

Darrin M. York is a Professor of Chemistry at Rutgers. He obtained his Ph.D. in chemistry at the University of North Carolina at Chapel Hill with Prof. Lee Pedersen and was a postdoctoral researcher at Duke with Prof. Weitao Yang and at Harvard and Université Louis Pasteur in Strasbourg with Prof. Martin Karplus.

■ REFERENCES

(1) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708−2728.

(2) Muller, R. P.; Warshel, A. Ab initio calculations of free energy barriers for chemical reactions in solution. *J. Phys. Chem.* **1995**, *99*, 17516−17524.

(3) Giese, T. J.; York, D. M. Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations. *J. Chem. Phys.* **2007**, *127*, No. 194101.

(4) Crawford, T. D. Coupled cluster calculations of vibrational raman optical activity spectra. *ChemPhysChem* **2011**, *12*, 3442−3448.

(5) Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Simulation of NMR data reveals that proteins local structures are stabilized by electronic polarization. *J. Am. Chem. Soc.* **2009**, *131*, 8636−8641.

(6) He, X.; Wang, B.; Merz, K. M., Jr. Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B* **2009**, *113*, 10380−10388.

(7) Zhu, T.; He, X.; Zhang, J. Z. H. Fragment density functional theory calculation of NMR chemical shifts for proteins with implicit solvation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7837−7845.

(8) Zanni, M. T.; Hochstrasser, R. M. Two-dimensional infrared spectroscopy: a promising new method for the time resolution of structures. *Curr. Opin. Struct. Biol.* **2001**, *11*, 516−522.

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

(10) 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.

(11) Jacobson, L. D.; Herbert, J. M. An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion. *J. Chem. Phys.* **2011**, *134*, No. 094118.

(12) Giese, T. J.; Chen, H.; 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.

(13) Duke, R. E.; Starovoytov, O. N.; Piquemal, J.; Cisneros, G. A. GEM*: A molecular electronic density-based force field for molecular dynamics simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1361−1365, DOI: 10.1021/ct500050p.

(14) Donchev, A. G.; Galkin, N. G.; Pereyaslavets, L. B.; Tarasov, V. I. Quantum mechanical polarizable force field (QMPFF3): Refinement and validation of the dispersion interaction for aromatic carbon. *J. Chem. Phys.* **2006**, *125*, No. 244107.

(15) Wollacott, A. M.; Merz, K. M., Jr. Assessment of semiempirical quantum mechanical methods for the evaluation of protein structures. *J. Chem. Theory Comput.* **2007**, *3*, 1609−1619.

(16) 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.

(17) Yang, W. Electron density as the basic variable: A divide-and-conquer approach to the ab initio computation of large molecules. *J. Mol. Struct.* **1992**, *255*, 461−479.

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

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

(20) 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.

(21) 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.

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

(23) Giese, T. J.; Chen, H.; 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.

(24) Gaus, M.; Goez, A.; Elstner, M. Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.

(25) Huang, M.; Giese, T. J.; Lee, T.-S.; York, D. M. Improvement of DNA and RNA sugar pucker profiles from semiempirical quantum methods. *J. Chem. Theory Comput.* **2014**, *10*, 1538−1545.

(26) Korth, M. Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields. *J. Chem. Theory Comput.* **2010**, *6*, 3808−3816.

(27) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J. Mol. Biol.* **2003**, *327*, 549−560.

(28) de Courcy, B.; Piquemal, J.; Garbay, C.; Gresh, N. Polarizable water molecules in ligand-macromolecule recognition. Impact on the relative affinities of competing pyrrolopyrimidine inhibitors for FAK kinase. *J. Am. Chem. Soc.* **2010**, *132*, 3312−3320.

(29) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290−6295.

(30) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Wang, L.; Whitfield, H. J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745−749.

(31) Reynols, C. H.; Holloway, M. K. Thermodynamics of ligand binding and efficiency. *ACS Med. Chem. Lett.* **2011**, *2*, 433−437.

(32) Ji, C. G.; Zhang, J. Z. H. Electronic polarization is important in stabilizing the native structures of proteins. *J. Phys. Chem. B* **2009**, *113*, 16059−16064.

(33) Urakubo, Y.; Ikura, T.; Ito, N. Crystal structural analysis of protein-protein interactions drastically destabilized by a single mutation. *Protein Sci.* **2008**, *17*, 1055−1065.

(34) Kobe, B.; Deisenhofer, J. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **1995**, *374*, 183−186.

(35) Kryger, G.; Harel, M.; Giles, K.; Toker, L.; Velan, B.; Lazar, A.; Kronman, C.; Barak, D.; Ariel, N.; Shafferman, A.; Silman, I.; Sussman, J. L. Structures of recombinant native and E202Q mutant human acetylcholinesterase complexed with the snake-venom toxin fasciculin-II. *Acta. Crystallogr.* **2000**, *D56*, 1385−1394.

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

(37) Jurecka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.

(38) Řežač, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(39) Bryantsev, V. S.; Diallo, M. S.; van Duin, A.; Goddard, W. A., III Evaluation of B3LYP, X3LYP, and M06-class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016−1026.

(40) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from X-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138*, No. 074506.