

## Topical Review

# Quantum mechanical force fields for condensed phase molecular simulations

Timothy J Giese and Darrin M York<sup>1</sup>

Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854-8087, United States of America

E-mail: [Darrin.York@rutgers.edu](mailto:Darrin.York@rutgers.edu)

Received 20 February 2017, revised 26 June 2017

Accepted for publication 28 June 2017

Published 17 August 2017

**Abstract**

Molecular simulations are powerful tools for providing atomic-level details into complex chemical and physical processes that occur in the condensed phase. For strongly interacting systems where quantum many-body effects are known to play an important role, density-functional methods are often used to provide the model with the potential energy used to drive dynamics. These methods, however, suffer from two major drawbacks. First, they are often too computationally intensive to practically apply to large systems over long time scales, limiting their scope of application. Second, there remain challenges for these models to obtain the necessary level of accuracy for weak non-bonded interactions to obtain quantitative accuracy for a wide range of condensed phase properties. Quantum mechanical force fields (QMFFs) provide a potential solution to both of these limitations. In this review, we address recent advances in the development of QMFFs for condensed phase simulations. In particular, we examine the development of QMFF models using both approximate and *ab initio* density-functional models, the treatment of short-ranged non-bonded and long-ranged electrostatic interactions, and stability issues in molecular dynamics calculations. Example calculations are provided for crystalline systems, liquid water, and ionic liquids. We conclude with a perspective for emerging challenges and future research directions.

Keywords: quantum mechanical force field, linear scaling, electronic structure

(Some figures may appear in colour only in the online journal)

**1. Introduction**

Molecular simulations can provide deep insight into chemical processes in the condensed phase by unveiling atomic-level details beyond that which can be directly measured experimentally. However, the predictive value of these calculations hinges on the accuracy of the models used to determine intra- and intermolecular forces, the time scales the simulations can access, and the degree to which the spatial extent of the model system *in silico* accurately reflects the real system *in vitro* or *in vivo*. These considerations are problem-dependent and

can vary widely—but universally there is a trade-off between accuracy and computational feasibility that must be appropriately balanced.

Pairwise additive ‘molecular mechanics’ (MM) force fields are among the simplest, and most efficient models and have seen widespread use in simulations of large systems over long time scales [1]. Several next-generation force fields have been developed that include explicit electronic polarization and other more realistic features in the molecular interactions [2–4]. While these force fields are quite promising, they are also less computationally efficient, and in many cases have not yet achieved the same application scope and level of parameter refinement as the more mature pairwise additive models.

<sup>1</sup> Author to whom any correspondence should be addressed.

Many chemical applications demand models that go beyond molecular mechanics and explicitly use a quantum mechanical framework [5, 6]. Quantum mechanical (QM) models provide more transferability between varying systems of interest and chemical environments than traditional MM force fields. Furthermore, QM models are inherently capable of modeling electronic polarization, charge transfer, chemical reactions, and spectra. The literature describing the development of modern quantum electronic structure methods is vast; however, for the system sizes and time scales required for condensed phase simulations, the scope of practical methods is quite limited. By far the most widely used quantum methods applied to molecular simulations are those based on single-determinant wave functions constructed from molecular orbitals built up from localized atomic orbital basis sets. For the purposes of the present review, we will restrict discussion to this class of quantum models, which includes both approximate and *ab initio* methods based on Hartree–Fock and density-functional theory [7].

The applicability of molecular orbital-based quantum mechanical methods to molecular dynamics (MD) simulation of large systems is severely hindered by:

1. the large number of required integrals to be performed,
2. the construction of globally orthonormal molecular orbitals, and
3. the quality of computed results from an affordable Hamiltonian.

There are QM algorithms that forego the explicit calculation of orthonormal orbitals; relying instead on properties of the density matrix. The development of density matrix minimization [8, 9] and density matrix subsystem density functional methods [10] continue to attract attention, but have not yet found widespread adoption relative to orbital-based algorithms.

The number of electron repulsion integrals (ERIs) required in *ab initio* methods formally scales  $O(N^4)$ , where  $N$  is the number of atoms in the system. For moderately sized systems, this scaling can be reduced to  $O(N^3)$  by computing the ERIs through Cholesky decomposition or the resolution-of-identity approximation [11]. Alternatively, some Hamiltonians have been proposed that reduce the computational cost (prefactor) of density functional theory (DFT) methods by using very small atomic orbital (AO) basis sets and supplementing the Hamiltonian with empirical corrections to improve the quality of the results [12]. Semiempirical Hamiltonians reduce the formal scaling of the ERIs to  $O(N^2)$  either through the neglect of diatomic differential overlap (NDDO) approximation [13], or by computing the electrostatics from small, atom-centered auxiliary basis functions, such as the Slater atomic charges used in the self-consistent charge density-function tight-binding models DFTB2 [14] and DFTB3 [15, 16]. The formal scaling of *ab initio* and semiempirical Coulomb integral evaluation can be reduced to  $O(N \log N)$  by approximating the long-range interactions in aperiodic systems with the fast multipole method [17, 18]. Linear-scaling evaluation of electrostatics in periodic systems remains an active area of research and is discussed in section 4.

The construction of globally orthonormal molecular orbitals is usually accomplished by solving the Roothaan–Hall equations:

$$\mathbf{F}^\sigma \cdot \mathbf{C}^\sigma = \mathbf{S} \cdot \mathbf{C}^\sigma \cdot \mathbf{E}^\sigma \quad (1)$$

where  $F_{\mu\nu}^\sigma = \partial E(\mathbf{R}, \mathbf{P}) / \partial P_{\mu\nu}^\sigma |_{\mathbf{R}}$  is the spin-resolved Fock (or Kohn–Sham) matrix,  $E(\mathbf{R}, \mathbf{P})$  is the total energy of the electronic structure method,  $P_{\mu\nu}^\sigma = \sum_i n_i^\sigma C_{\mu i}^\sigma C_{\nu i}^\sigma$  is the AO representation of the spin-resolved density matrix,  $\mathbf{C}^\sigma$  is the matrix of spin-resolved MO coefficients,  $n_i^\sigma$  is a spin-resolved occupation number of MO  $i$ ,  $\mathbf{E}^\sigma$  is the diagonal matrix of spin-resolved MO eigenvalues, and  $S_{\mu\nu} = \int \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d^3r$  is the AO overlap matrix. Equation (1) is a generalized eigenvalue equation, whose solution formally scales  $O(N^3)$ . This bottleneck can be overcome by fragmenting the system into parts, such that each fragment independently solves equation (1) to obtain MOs that are locally (but not globally) orthonormal. The differences between the myriad number of fragment-based electronic structure methods occurs within their technical details that define how the fragments couple to one another. These differences are discussed in section 3.

The computational effort of ‘linear-scaling quantum mechanical’ (LSQM) methods [19, 20], as the name suggests, scales linearly with system size. Traditional LSQMs seek to systematically reproduce the results of a fully-coupled QM calculation without introducing empirically-fit parametric corrections to the interaction energy, instead preferring to explicitly model the inter-fragment MO coupling through a buffer region or many-body expansion. Although LSQM methods scale desirably with respect to system size, they are substantially more expensive than QM/MM calculations, and they may offer significant computational advantages relative to a standard QM method only when large systems are considered. Quantum mechanical force fields [21, 22] (QMFFs) are LSQM methods that replace elements of theoretical rigor by introducing parametrically fitted functions to the interaction energy (beyond those that may be contained in the base QM method). The parametrically fitted corrections are chosen to minimize the use of explicit inter-fragment MO coupling. Consequently, QMFFs achieve higher computational savings than traditional LSQMs, and they are often faster than fully-coupled QM methods, even for small systems. Furthermore, the parametric tuning of the interactions provides a means for improving the accuracy of QMFFs for certain properties.

Although QMFFs introduce empirical functions to improve upon inter-fragment interactions, the quality of their intra-fragment interactions continues to rely upon the approximations within the underlying QM Hamiltonian. Recent advances in semiempirical and DFT methods are briefly discussed in section 2. QMFFs are methods that attempt to leverage the benefits provided by LSQMs and combined quantum mechanical/molecular mechanical (QM/MM) [6, 23] approaches to produce a fully QM method that is able to simultaneously achieve high accuracy and efficiency. The first QMFF suited for condensed phase simulation applications was Gao’s ‘molecular orbital derived empirical potential for liquid simulations’ (MODEL) method; [24] however, the term ‘quantum mechanical force field’ has been used to describe this type of method

only recently [21, 22]. Within the current context, QMFFs are fundamentally built on a QM framework rather than simply using quantum mechanical results to fit an empirical molecular mechanical potential, as has been used successfully in the derivation of class II [25] and other classical [26, 27] force fields. It should be pointed out that strategies for developing quantum methods for large molecular systems differ in some ways from those for solid state systems owing to the need to address distinct challenges. Solid state applications often involve strongly correlated systems characterized by networks of bonds that require specialized embedding methods [28–34] to model accurately [35]. Molecular systems, which are the focus of this topical review, interact largely through weaker non-bonding interactions. Consequently, molecular systems typically require more phase space sampling to achieve reliable condensed phase properties.

Recent applications of QMFFs and LSQMs to MD simulation, drug docking, chemical reaction profiles, electron transfer reactions, NMR chemical shifts, and IR, Raman, and vibrational circular dichroism spectra are discussed in section 6.

## 2. Intra-fragment QM Hamiltonians

QMFFs and LSQMs require the choice of a QM Hamiltonian through a suitable compromise between performance and the Hamiltonian's accuracy for the intended application. Most MD applications, for example, require a relatively large number of particles and a significant amount of phase-space sampling to make a meaningful comparison with experiment. Semiempirical Hamiltonian QMFFs offer the greatest promise for this purpose, at least at the present time. There are two popular classes of semiempirical models in widespread use: modified neglect of diatomic overlap (MNDO) models [13, 36, 37] and density functional tight binding (DFTB) models [38, 39]. There are advantages and disadvantages for both types of Hamiltonian models [40, 41]. Most MNDO-based models improperly model torsion potentials and nucleobase sugar puckering due to the lack of orthogonalization effects resulting from the NDDO approximation [42]. Orthogonalization corrections (e.g. the OMx methods [43, 44]) have been introduced within the NDDO framework, and have been demonstrated to be a promising improvement [45], although reported applications to molecular dynamics simulations are scarce. DFTB-based models do not make the NDDO approximation, but—unlike MNDO models—the DFTB electrostatics are modeled solely through atomic charges, without regard to higher-order multipoles. MNDO and DFTB models both suffer from the limitations resulting from the use of minimal valence basis sets [46] and the range of their applicability caused by their limited training and inherent transferability of the model parameters. These limitations have motivated the reparametrization of models for general use or for system-specific applications [15, 16, 47–51], and the development of MNDO Hamiltonians which include orthogonalization corrections [45] and improved polarizabilities [52–54]. DFTB methods have similarly been developed to improve polarizabilities

[55–57] and hydrogen bond strengths [58]. Exploratory DFTB methods have improved nonbond interactions by incorporating multipolar electrostatics [59–63], and DFTB pairwise potentials continue to be parametrized to widen the scope of applications [64–69].

The calculation of infrared (IR) and Raman spectra is challenging for semiempirical Hamiltonians due to their limited AO basis sets [70]. Therefore, recourse to higher-level *ab initio* DFT methods are needed for these applications. Recently, there has been growing interest in the development of so-called 'range-separated' density functionals, which have shown promise for making improved prediction of x-ray adsorption modeling [71], optoelectronic properties [72], noncovalent interactions [73], and sulfur chemistry [74], but which may offer only limited improvement of transition metal reactions in comparison to standard hybrid functionals [75].

An important consideration for both semiempirical and *ab initio* Hamiltonians is the treatment of dispersion [76]. For example, the use of dispersion-corrected *ab initio* density functionals [77] and semiempirical density functional tight binding Hamiltonians [78] has been shown to improve crystal structure prediction [79, 80]. However, if the fragments are individual molecules, then QMFFs can couple the fragments with an empirical dispersion model even though the underlying QM Hamiltonian's treatment may be inadequate [81].

## 3. Inter-fragment coupling schemes

Many algorithms founded upon fragmentation have been developed, classified, and previously discussed [21, 22, 82–92]. In this section, we emphasize the differences between fragment algorithms through the discussion of four types of methods:

1. The 'divide-and-conquer' (DC) LSQM method.
2. The fragment molecular orbital (FMO) LSQM method.
3. Force fields founded upon underlying QM calculations, such as the effective fragment potential (EFP), the sum of interactions between fragments *ab initio* (SIBFA), and Gaussian electrostatic model (GEM) methods.
4. The eXplicit polarization (X-Pol) and modified divide-and-conquer (mDC) QMFF methods.

The DC method [93–101] is a LSQM approach that has seen successful use in the simulation of liquid water [102], the prediction of protein  $pK_a$  values [103], and the calculation of NMR chemical shifts [104, 105]. In this scheme, the system is fragmented into distinct parts, and a fragment's neighbors are considered as a 'buffer' region when constructing the fragment MOs. A Fock matrix is constructed for each fragment (including its buffer), whose diagonalization produces a set of locally orthonormal MOs. By itself, this procedure would double-count the electron density between fragments; thus, the density matrix of each fragment+buffer region is partitioned such that:

- If both AOs belong to the buffer, then those density matrix elements are removed.

- If both AOs belong to the fragment, then the density matrix elements remain unaltered.
- If an AO pair is split between the buffer and the fragment, then the density matrix elements are halved.

The use of a buffer region within the DC method causes the fragment MOs to directly couple to their neighbors, but the nonzero Fock matrix elements are influenced by the electrostatics of the entire system. Furthermore, additional *ad hoc*, empirical potentials are not introduced; that is, the non-electrostatic inter-fragment coupling occurs solely through the inclusion of a buffer. The method is made more accurate by increasing the size of the buffer.

The FMO method is considered here as a LSQM [106–108] owing to its close connection to the underlying *ab initio* framework. Analytic gradients for the FMO method have only recently been derived [106], enabling its application in MD simulation [109, 110]. This method fragments the system into nonoverlapping components. The MOs of each system are optimized in the electrostatic environment of the other fragments. Because the fragment MOs give rise to the electrostatic environment, the procedure is iterated until both the fragment MOs and the electrostatic environment reach self-consistency. Upon convergence, the energetic effects of inter-fragment MO coupling is modeled by a many-body expansion. Specifically, the many-body corrections are computed by self-consistently solving for the MOs of a fragment cluster in the presence of the fixed electrostatic potential of the remaining monomers. In the FMO2 method, for example, the inter-fragment coupling energy of fragment-pairs are corrected in the static field of the remaining  $N - 2$  fragments. Similarly, the FMO3 method corrects three-body MO couplings. The method is made more accurate by increasing the order of the many-body corrections. A Variational Many-Body expansion (VMB) method has been introduced by Gao and Wang [111] that extends the basic idea of FMO by variationally optimizing each term in the many-body expansion, as opposed to performing single-point energy corrections in the static field of the remaining fragment monomers.

The EFP [112–114], SIBFA [115–119], and GEM [115, 117, 120–123] methods are substantially more similar to classical, polarizable force fields, because the the inter-fragment interactions are computed from empirical functions; however, many—if not all—of the parameters within those empirical functions are produced from the analysis of an underlying QM calculation. For example, the inter-fragment electrostatics are computed from auxiliary Gaussian multipoles or damped, distributed point-multipoles arising from a reference *ab initio* charge density. Similarly, the fragments polarize to one another through linear-response dipoles, whose polarizability tensor can be constructed from coupled-perturbed Hartree–Fock/Kohn–Sham theory. The *ab initio* polarizability tensors can be used to generate  $C_6$  dispersion coefficients to model London forces with Tang and Toennies [124] damped dispersion potentials. And the short-range, nonclassical, nonbonded repulsion is modeled either by inter-fragment electron density overlap or as a function of MO overlap [116, 125, 126]. By using cost-effective, empirical functional forms, these

methods do not intend to formally reproduce the results of a fully-coupled MO *ab initio* calculation. Instead, the fragment *ab initio* calculations are used to nonempirically determine the parameters (or substantially reduce the amount of empirical fitting) appearing within the functional forms.

The X-Pol [24, 127–132] and mDC [21, 59–61, 81] methods are QMFFs in which the system is decomposed into nonoverlapping fragments, the fragment MOs are coupled to the other fragments only through electrostatics, and the fragment MOs of all systems are variationally optimized. Specifically, each fragment diagonalizes its own Fock/Kohn–Sham matrix  $\mathbf{F}^{\sigma,A}$ , where  $A$  indexes the fragment. The diagonalization of  $\mathbf{F}^{\sigma,A}$  produces a set of fragment-localized MO coefficients,  $\mathbf{C}^{\sigma,A}$ , which define a fragment-localized density matrix  $P_{\mu\nu}^A = \sum_{\sigma,i} n_i^{\sigma,A} C_{\mu i}^{\sigma,A} C_{\nu i}^{\sigma,A}$ . The electrostatics of the nuclei and electron density can be modeled from atomic point multipole expansions,  $\mathbf{q}^A$ , to efficiently compute the Coulomb interactions between fragments. Unlike the FMO method, the inter-fragment MO coupling is not corrected via a many-body expansion. Instead, X-Pol and mDC model the energetic effect of the inter-fragment MO coupling with empirical functions. Typically, the empirical correction for nonbonded interactions is a standard Lennard-Jones potential; however, more advanced charge-dependent van der Waals models have also been used [81]. The X-Pol energy is:

$$E_{\text{X-Pol}} = \sum_A E_{\text{QM}}^A(\mathbf{P}^A) + \frac{1}{2} \sum_{B>A} E_{\text{QM/MM}}^{AB}(\mathbf{P}^A, \mathbf{q}^B) + E_{\text{QM/MM}}^{BA}(\mathbf{P}^B, \mathbf{q}^A) \quad (2)$$

where

$$E_{\text{QM/MM}}^{AB}(\mathbf{P}^A, \mathbf{q}^B) = \sum_{\substack{a \in A \\ b \in B}} \frac{Z_a q_b}{R_{ab}} - \sum_{\substack{\mu \in A \\ b \in B}} P_{\mu\nu}^A q_b \int \frac{\chi_{\mu}(\mathbf{r}) \chi_{\nu}(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_b|} d^3r + E_{\text{vdW}}(\mathbf{R}^A, \mathbf{R}^B) \quad (3)$$

is the QM/MM interaction energy between fragments  $A$  and  $B$ , in which fragment  $A$  is treated as the QM region that interacts with the atomic charges in fragment  $B$ , and  $E_{\text{vdW}}$  is the van der Waals energy. The mDC energy is similarly defined:

$$E_{\text{mDC}} = \sum_A E_{\text{QM}}^A(\mathbf{P}^A) + \sum_{B>A} E_{\text{inter}}(\mathbf{q}^A, \mathbf{q}^B) \quad (4)$$

where the interaction energy is computed directly from the atomic multipole expansions and van der Waals potential.

$$E_{\text{inter}}(\mathbf{q}^A, \mathbf{q}^B) = \sum_{\substack{l\mu \in A \\ j\kappa \in B}} q_{l\mu}^A q_{j\kappa}^B \frac{C_{l\mu}(\nabla_a)}{(2l-1)!!} \frac{C_{j\kappa}(\nabla_b)}{(2j-1)!!} \frac{1}{R_{ab}} + E_{\text{vdW}}(\mathbf{R}^A, \mathbf{R}^B) \quad (5)$$

$C_{l\mu}(\nabla_a)$  is a real-valued spherical tensor gradient operator (STGO) acting on the coordinates of atom  $a$ . A STGO is a solid harmonic function, whose Cartesian coordinate arguments have been replaced by Cartesian gradient operators. At the present time, the mDC method has only been used in conjunction with the DFTB3 semiempirical Hamiltonian [15, 16, 39]. A standard DFTB3 calculation computes the long-range electrostatics via an auxiliary Slater monopole representation of the density, and the multipolar character of the interactions is only treated in the short-range regime within the tight-binding matrix. Therefore, the use of DFTB3 within the X-Pol formalism would limit the inter-fragment Coulomb

interactions to those between atomic charges, and the explicit multipolar interactions between fragments would be completely neglected because of the fragment-localized MOs. Consequently, it has been shown that water clusters lose their hydrogen bond angles, instead choosing to adopt TIP3P-like [133], planar hydrogen bonding motifs [21, 59, 61]. To correct for this deficiency, the mDC method concocts a new, auxiliary atomic multipole representation of the fragment density that is solely used for the inter-fragment Coulomb interactions.

When the fragments sever covalent bonds, X-Pol makes use of the Generalized Hybrid Orbital QM/MM method [134–139], and mDC makes use of the link-atom approach [140–143], although other capping strategies have been applied in similar QMFFs [144–146]. Ideas have been proposed for nonempirically incorporating short-range non-bond repulsions by introducing inter-fragment orthogonality restraints into X-Pol [147] but practical applications of those ideas have not yet been explored.

It is worth mentioning that an alternative approach to QMFF development based on more reactive force fields that employ purely empirical potentials has been valuable. Most notably is the ReaxFF [148] framework that uses classical interatomic potentials to model chemical reactions [149–153] in large, complex systems. The ReaxFF does not attempt to solve explicitly for the electronic structure, and so is different conceptually from the methods described in the present topical review, but nonetheless is sufficiently computationally efficient to be applied in very large-scale molecular simulations [154].

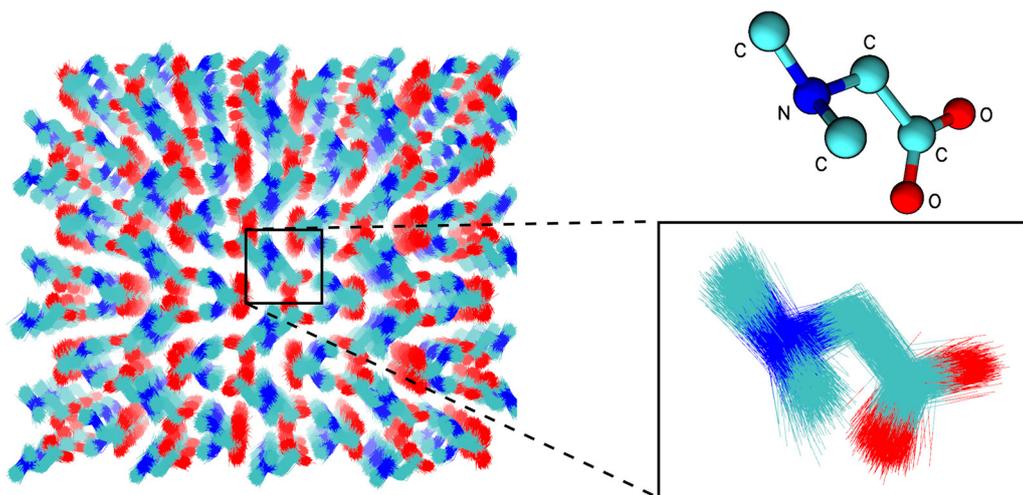
#### 4. Long-ranged electrostatic interactions

The primary source of inter-fragment MO coupling within QMFFs occurs indirectly through the Coulomb interactions between the fragments, and the manner in which the interactions are performed is thus important. For example, applications of the mDC method found that the inclusion of atomic dipoles and quadrupoles were paramount for obtaining proper hydrogen bond angles in water clusters [21, 59, 61]. Furthermore, condensed phase simulations of QMFFs require a treatment for the long-range electrostatic interactions, such as Ewald's method [155]. Ewald's method has seen widespread adoption within standard MM software packages after the development of particle mesh Ewald (PME) [156–159] because electrostatic cutoff methods have been found to produce structural artifacts in biomolecules [160–162] and in the simulations of thermodynamic properties of water [163–166]. Because the PME method was originally designed for static atomic charges, the literature has witnessed a recent resurgence in interest toward the development of long-range electrostatic models for atomic multipoles [167–169], semiempirical Hamiltonians [170, 171], an *ab initio* calculations [172–177].

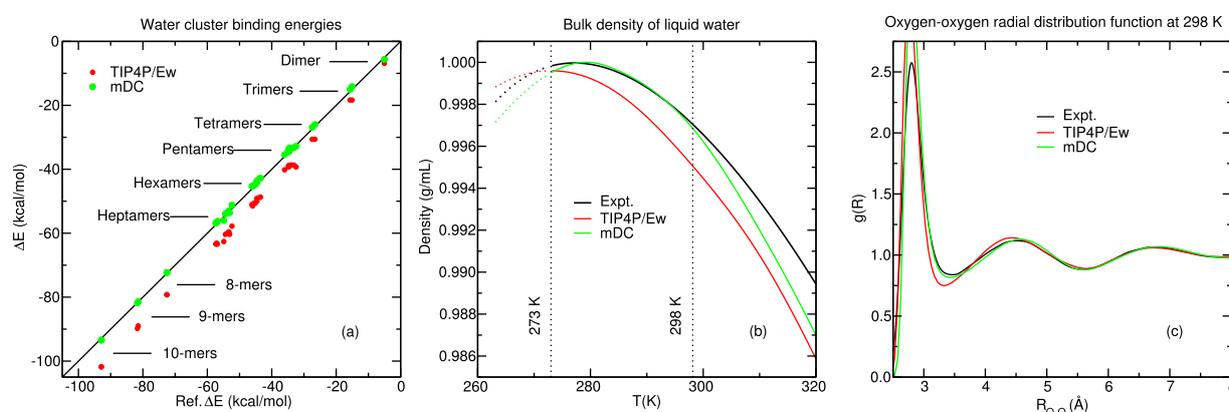
The usage of MNDO models within the X-Pol QMFF naturally affords the adaptation of the semiempirical PME method [171] originally developed for QM/MM calculations [170, 178, 179], whereas the use of high-order atomic multipoles within

the mDC model motivated the development of a multipolar PME method [60, 167]. It was found that the inclusion of atomic dipoles and quadrupoles slowed the mDC calculations by only a factor of 1.5 relative to a charge-only model [167]. In contrast, the semiempirical QM/MM PME method [170, 178, 179] has been found to *not* be readily amenable to *ab initio* Hamiltonians because the use of simple charge-mapping procedures, such as Mulliken or Löwdin charge analysis, causes SCF convergence instabilities or spurious electron polarizations when non-minimal basis sets are used [173, 174]. In order to overcome this problem, some have resorted to using CHELPG [180] charge-fitting procedures specifically to avoid the development of a wholly new Ewald method [173, 174, 181–185]. Others groups have opted to simply perform real-space electrostatics within a large (22 Å) cutoff that smoothly dampens the electrostatic interactions [175, 176], or to represent the long-range electrostatics by choosing a set of charges on a discretized sphere [177], thereby eliminating (or reducing) the necessary changes to the QM software package. Another approach to avoid using Ewald's method with *ab initio* Hamiltonians include the reaction field method [186, 187]; however, the predominant choice continues to be the use of real-space electrostatic truncation [127, 146, 188–194].

Very recently, a new type of *ab initio* QM/MM PME method—the ambient-potential composite Ewald (CEw) method—has been introduced [172]. The CEw method allows the *ab initio* nuclei and electron density to interact with the environment without resorting to an atomic charge representation of the QM density. Furthermore, the method does not require more plane waves than what are used in traditional MM calculations. Explicit calculation of the AO-product Fourier coefficients are avoided by numerically integrating the reciprocal-space Ewald potential via the molecular quadrature grid used to evaluate the DFT exchange-correlation potential. Additionally, the CEw method models the QM region's periodic replicas as static MM point charges, so that the Ewald contribution to the Fock matrix can be performed once, before the SCF procedure begins. In this sense, the evaluation of the long-range electrostatics within the CEw method becomes analogous to a pre-SCF evaluation of a local density approximation DFT exchange-correlation potential. In that work [172], it was argued that the cost associated of the underlying *ab initio* method wouldn't be fully taken advantage of if the electrostatics were to be performed merely through atomic charges. Furthermore, the CEw method circumvents the SCF instabilities and spurious polarizations that occur when simple charge-mapping procedures are used. Finally, the authors found that potential of mean force (PMF) profiles and radial distribution functions were sensitive to the treatment of long-range electrostatics. Specifically, electrostatic truncation and switching can cause artificial solvation shells to appear near the electrostatic cutoff, which changes the free energy difference as a dianionic solute dissociates into two monoanionic components. The CEw method is immediately amenable to an *ab initio* VMB fragment method in which the monomer fragments were modeled by a MM force field.



**Figure 1.** Atomic fluctuations of a *n,n*-dimethylglycine crystal from 220 ps of NVT simulation at 275 K using the mDC QMFF.



**Figure 2.** (a) Comparison of water cluster interaction energies to high-level *ab initio* results (complete basis set extrapolated RI-MP2 with CCSD(T) corrections). Reference energies taken from [219]. (b) Comparison of liquid water bulk densities. Experimental values taken from [220] and [221]. (c) Condensed phase water pair distribution functions. Experimental values taken from [222].

## 5. Illustrative examples

To illustrate the computational performance of *ab initio* density-functional models within fragment-based calculations, we consider two systems: metenkephalin, and a UpA dinucleotide (U5A3), shown in figure 3, to examine the performance and resource efficiency as the solute molecule is fragmented into smaller parts.

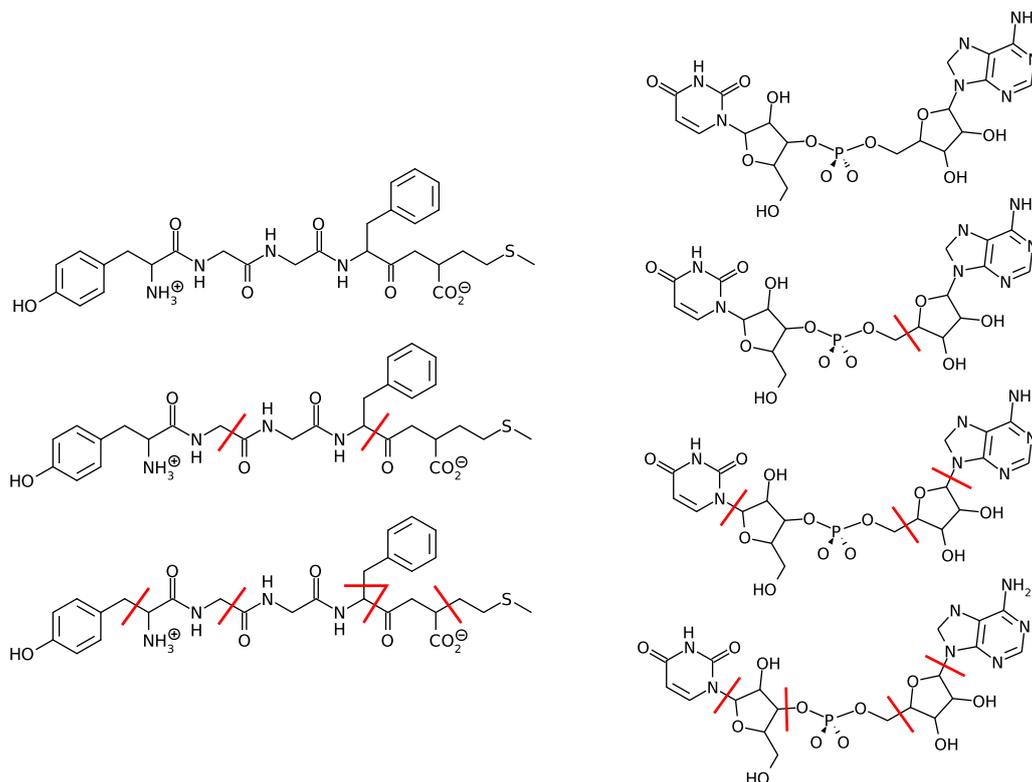
The metenkephalin solute is solvated in a  $42^3 \text{ \AA}^3$  box of TIP4P/Ew waters. The U5A3 system is solvated in a  $46^3 \text{ \AA}^3$  box of TIP4P/Ew waters and an ion environment of 7  $\text{Na}^+$  and 6  $\text{Cl}^-$ . These systems were prepared by performing pure-MM simulations (1 fs/step) for 1 ns at 298 K and 1 atm to equilibrate the density. This was followed by 5 ps of NVT equilibration using QM/MM (1 fragment), where the solute was treated quantum mechanically with PBE0/6-31G\*. The equilibrated systems were then used to compare the fragment calculations described below.

The total potential energy of the fragment-based calculations is given by equation (6):

$$E(\mathbf{R}) = E_{\text{MM}}(\mathbf{R}) + \sum_A [E_{\text{QM/MM}}^A(\mathbf{R}; \mathbf{P}^A) - E_{\text{MM}}(\mathbf{R})] \quad (6)$$

In other words, the system's MM energy is modified by computing QM/MM corrections for each fragment. A fragment's energy correction is computed by performing a QM/MM energy calculation, in which the fragment is the QM region and the remainder of the system is MM. The correction thus replaces the fragment's MM self-energy with its QM self-energy and further replaces the fragment's MM interactions with the remainder of the system with the corresponding QM/MM interactions. Equation (6) can be viewed as being analogous to a first-order FMO method; however, the fragments polarize to the MM electrostatic field rather than the field of fixed charge densities resulting from independent gas-phase QM calculations. Equation (6) differs from X-Pol, mDC, and higher-order FMO methods in that the fragment calculations are decoupled from each other. Therefore, the series of QM/MM SCF procedures are independent of one another. We compute all QM/MM calculations in tandem on different CPU cores, and the number of cores assigned to each QM/MM calculation is empirically adjusted so that each calculation completes in approximately the same amount of time.

Table 1 provides analysis of NVT simulation timings using equation (6) with PBE0/6-31G\* (and the CEw method for long-range electrostatics [172]) for the fragmentation schemes



**Figure 3.** The fragmentation schemes used for metenkephalin (left column) and U5A3 (right column). The top row is a single fragment. The red lines indicate the QM/MM boundaries. The metenkephalin is fragmented into 1, 3, and 6 fragments. The U5A3 system is fragmented into 1, 2, 4, and 5 fragments.

**Table 1.** PBE0/6-31G\* simulation performance and resource efficiency of the metenkephalin and U5A3 systems as a function of fragmentation (see figure 3).

$N_{\text{frag}}$	Metenkephalin			U5A3			
	1	3	6	1	2	4	5
$N_{\text{QM}}/N_{\text{CPU}}$	75/96	29/96 29/96 21/95 0/1	16/96 16/96 15/96 15/96 12/96 11/95 0/1	62/96	36/144 28/95 0/1	26/132 15/60 15/24 12/23 0/1	18/120 15/96 15/60 12/59 10/48 0/1
sec/step	20.1	3.8	0.5	18.5	6.9	1.6	0.6
ps $\text{d}^{-1}$	4.3	23.0	168.1	4.7	12.5	54.9	154.1
SU $\text{ps}^{-1}$	535.8	300.5	82.2	490.2	460.8	104.9	59.8

shown in figure 3. The  $N_{\text{QM}}/N_{\text{CPU}}$  notation specifies the number of QM + link atoms and the number of CPU cores used for a fragment. The ‘0/1’ entries are pure-MM calculations performed on a single core. The row marked ‘ps  $\text{d}^{-1}$ ’ is the amount of simulation achieved in one day using a 1 fs time step. Larger values of ps  $\text{d}^{-1}$  indicate greater simulation performance. The row marked ‘SU  $\text{ps}^{-1}$ ’ are the number of service units (core-hours) required to perform 1 ps of simulation. Smaller values of SU  $\text{ps}^{-1}$  indicate a more efficient use of the compute resources.

With 96 cores, each time step takes 20.1 s for the full metenkephalin molecule. This time is reduced by a factor of 40 when using the highest fragmentation scheme (6 fragments).

The increase in throughput would allow up to 168 ps of simulation per day. Similar timing results are apparent for the U5A3 system. Clearly, the fragmentation scheme greatly influences advantage of the fragmentation the computational efficiency: the smaller the fragments, the less Fock matrix elements that need to be constructed, and the smaller the matrices that need to be diagonalized.

The fragmentation schemes shown in figure 3 are different models for the total energy, and they will each yield their own set of atomic forces. Table 2 compares the extent to which fragmentation alters the net forces. Specifically, the tabulated values are time-averaged relative force differences (RFDs):

**Table 2.** Relative force differences (equation (7)) between the PBE0/6-31G\* fragmentation schemes (see table 1) and various reference Hamiltonians. The reference force evaluations do not fragment the QM region. The force differences between the MM force field and the references are shown for comparison.

$\mathbf{f}_{\text{ref}}$	Metenkephalin				U5A3				
	1	3	6	MM	1	2	4	5	MM
All-atom relative force differences									
$\mathbf{f}_{\text{PBE0/6-31G*}}$	0.00	0.05	0.05	0.05	0.00	0.01	0.03	0.03	0.06
$\mathbf{f}_{\text{PBE/6-31G*}}$	0.03	0.06	0.05	0.06	0.03	0.04	0.05	0.05	0.08
$\mathbf{f}_{\text{PBE0/6-311G**}}$	0.01	0.05	0.05	0.05	0.01	0.02	0.03	0.04	0.06
Solute atom relative force differences									
$\mathbf{f}_{\text{PBE0/6-31G*}}$	0.00	0.26	0.28	0.49	0.00	0.11	0.25	0.29	0.56
$\mathbf{f}_{\text{PBE/6-31G*}}$	0.27	0.38	0.41	0.60	0.30	0.32	0.40	0.42	0.69
$\mathbf{f}_{\text{PBE0/6-311G**}}$	0.09	0.27	0.29	0.48	0.11	0.16	0.28	0.32	0.56
Non-boundary solute atom relative force differences									
$\mathbf{f}_{\text{PBE0/6-31G*}}$	0.00	0.21	0.22	0.49	0.00	0.08	0.15	0.20	0.56
$\mathbf{f}_{\text{PBE/6-31G*}}$	0.27	0.35	0.37	0.60	0.30	0.31	0.35	0.36	0.69
$\mathbf{f}_{\text{PBE0/6-311G**}}$	0.09	0.23	0.24	0.48	0.11	0.13	0.19	0.23	0.56
Non-boundary, non-adjacent solute atom relative force differences									
$\mathbf{f}_{\text{PBE0/6-31G*}}$	0.00	0.17	0.16	0.49	0.00	0.06	0.09	0.12	0.56
$\mathbf{f}_{\text{PBE/6-31G*}}$	0.27	0.27	0.30	0.60	0.30	0.31	0.35	0.37	0.69
$\mathbf{f}_{\text{PBE0/6-311G**}}$	0.09	0.21	0.22	0.48	0.11	0.13	0.16	0.19	0.56

$$\text{RFD} = \frac{1}{N_{\text{frames}}} \sum_{t=1}^{N_{\text{frames}}} \sqrt{\frac{|\mathbf{f}_t - \mathbf{f}_{\text{ref},t}|^2}{|\mathbf{f}_{\text{ref},t}|^2}} \quad (7)$$

where  $\mathbf{f}_t$  is the array of net atomic forces evaluated at frame  $t$  of a reference trajectory consisting of  $N_{\text{frames}}$  frames, and  $\mathbf{f}_{\text{ref},t}$  is a corresponding array of forces evaluated at the same set of coordinates using a reference Hamiltonian. The reference trajectories consist of 1000 frames taken from a 100ps NVT simulation performed at the highest level of fragmentation (6 for metenkephalin and 5 for U5A3). The forces were obtained by re-evaluating each Hamiltonian at every frame. The ‘all-atom’ RFDs include the forces from the solute and solvent atoms; that is, the length of the force arrays include all atoms in the system. The ‘solute atom’ RFDs are computed from a petite list of forces that excludes the MM solvent. The ‘non-boundary solute atom’ RFDs further excludes the solute atom pairs whose bond is severed by the fragmentation. The ‘non-boundary, non-adjacent solute atom’ RFDs exclude the fragment boundary atoms and those atoms covalently bonded to the boundary atoms. The columns labeled ‘MM’ are the forces generated from the Amber ff14SB force field [195], which have been included for comparison.

As expected, fragmentation of the solute changes the forces, and the extent of the change increases as the solute is carved into smaller fragments. The solute atom RFDs suggest that the changes incurred by fragmenting the PBE0/6-31G\* solute are similar in magnitude to the changes observed upon switching to the PBE/6-31G\* Hamiltonian *without* fragmentation. The contribution of fragmentation to the solute atom RFDs are generally larger than having increased the size of the AO basis to PBE0/6-311G\*\* *without* fragmentation. Furthermore, the fragment boundary atoms and the atoms to which they covalently bonded contribute disproportionately to the solute atom

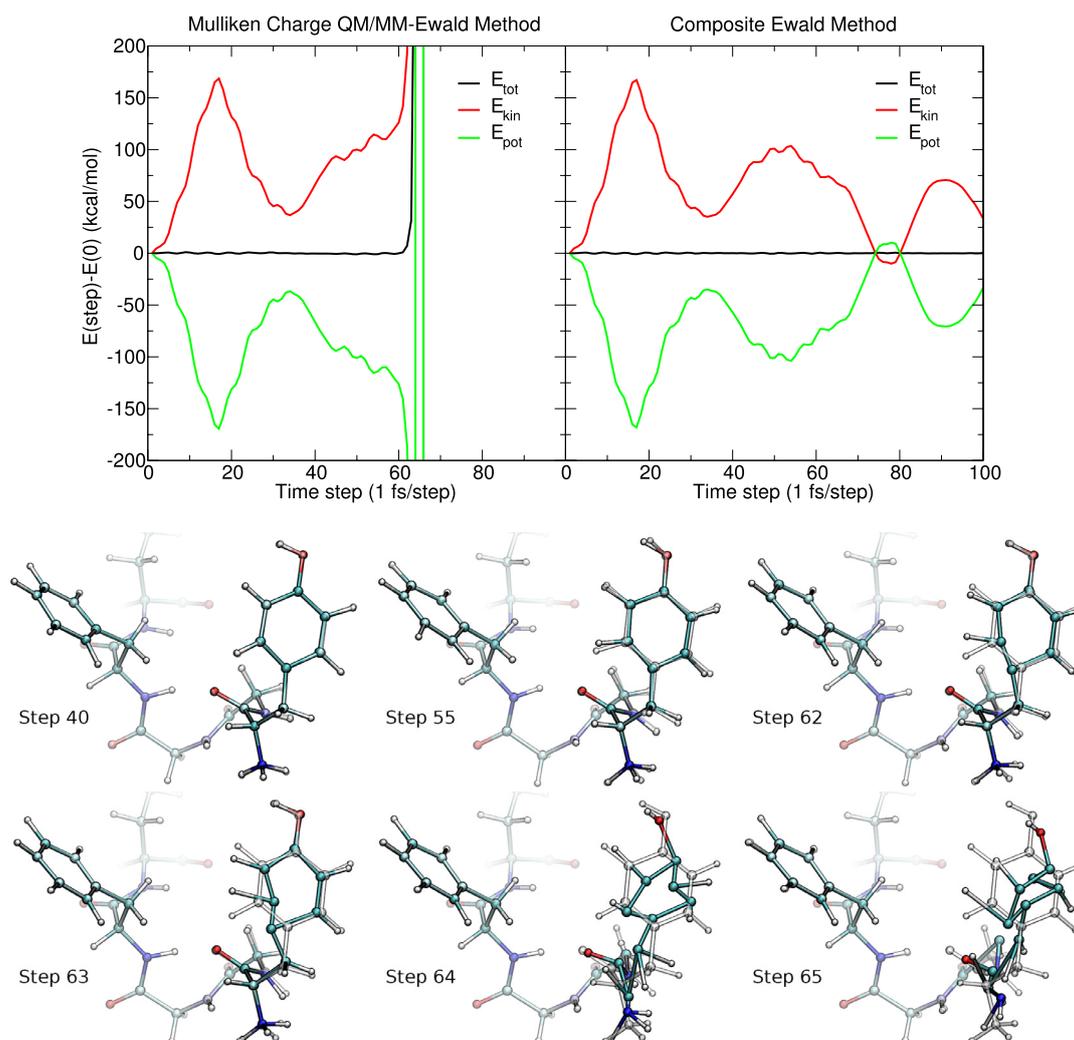
RFDs. Therefore, the solute atom RFDs could be significantly improved by specifically parametrizing the MM bond, angle, and torsion potentials that cross the fragmented bond. Table 2 evaluated these nonelectrostatic MM correction terms using the Amber ff14SB force field without specific reparametrization for this study.

To illustrate how long-range electrostatics affects the stability of fragment calculations, Figure 4 uses the metenkephalin peptide system by examining two NVE simulation trajectories that differ in how the Ewald summation is performed. Both trajectories are begun from the same initial conditions and both are evaluated using equation (6) and the 3-fragment scheme shown in figure 3 with the PBE0/6-311G\*\* QM Hamiltonian. The only difference between the two trajectories is the use of either the Mulliken-charge QM/MM-Ewald method [170, 178, 179] or the Composite Ewald method [172] for modeling the electrostatics.

Figure 4 emphasizes a subtle problem that can arise in molecular dynamics simulations using QMFFs in the condensed phase with Ewald electrostatics. If charge or multipolar mapping procedures are used to model electrostatic interactions between fragments, these mapping procedures may become vulnerable to instabilities under certain circumstances, such as with large basis sets (figure 4). The result is that simulations can become unstable and fail, unless special methods—such as the CEw method—are used.

## 6. Survey of applications

Although LSQM and QMFF methods continue to be rapidly developed, a wide range of applications has already been reported. Whereas LSQMs have been useful for many problems, the efficiency, and in some cases the accuracy of QMFFs have the potential to considerably extend the range



**Figure 4.** Energy profiles and selected snapshots from NVE simulations of the metenkephalin peptide system evaluated using the Mulliken-charge QM/MM-Ewald and Composite Ewald methods for long-range electrostatics.  $E_{\text{tot}}$ ,  $E_{\text{kin}}$ ,  $E_{\text{pot}}$  are the changes in the total, kinetic, and potential energy relative to the first step. The colored molecules are from the Mulliken-charge QM/MM-Ewald simulation. The black-and-white molecules are from the Composite Ewald simulation.

of applications, particularly in condensed phase molecular simulations. Fragment-based methods have seen recent use within molecular dynamics simulations [59, 60, 109, 146, 196, 197], and to explore the importance of many-body interactions [53, 198]. The FMO and mDC methods have been used in drug docking applications of cyclin-dependent kinase 2 inhibitor [59, 199], and to perform PMF profiles of chemical reactions [60, 200]. Proof-of-principle demonstrations have been performed in the development of models that afford electron transfer between fragments [201, 202]. There has been growing interest in using fragment methods—especially within the context of large biomolecules, such as the protein crambin, and crystal polymorphs of small drug-like compounds—for predicting spectra, including: IR [203], Raman [204], vibrational circular dichroism [205], and NMR chemical shifts [206–209]. Electrostatically embedded generalized molecular fractionation with conjugate caps (EE-GMFCC) method [85] has been applied to a number of macromolecular systems, including simulations of proteins [146], and calculations of RNA [210] and ion-water clusters

[211]. Molecular simulations using the frozen domain formulation of the FMO method have been performed to examine  $S_N2$  reactions in explicit solvent and protein-ligand binding free energies for the Trp cage protein [110]. ONIOM-based fragmentation algorithms have been applied within a hybrid extended-Lagrangian, post-Hartree–Fock Born–Oppenheimer *ab initio* molecular dynamics scheme to study protonated water clusters and polypeptide fragments [196]. FMO methods have been further applied using tight-binding models in nonadiabatic dynamics simulations to model charge transfer in the subphthalocyanine(SubPc)/ $C_{60}$  heterojunction [212].

The majority of QMFF applications to condensed phase simulations reported to date have utilized fast, approximate QM methods for construction of the intra-fragment QM Hamiltonian. The tremendous performance advantages afforded by these methods enable simulations to be performed on realistic-sized model systems of sufficient size and over time scales on time scales that yield meaningful statistical properties to be measured.

### 6.1. Water clusters and liquid water

Bulk water is the most common benchmark system for study in the liquid phase. It is well known that conventional force field models have difficulty in simultaneously describing water in environments that are outside the scope of their parameterization, which often are restricted to bulk water properties. For example, the TIP4PEw model for water [164] performs exceptionally well for bulk properties of water, such as the oxygen–oxygen pair distribution function, especially considering the extremely simplistic form of the MM potential. However, this model does not perform well for intermolecular interactions in small water clusters [21, 60] (figure 2). The mDC model, however, not only performs equally well as TIP4PEw for the pair distribution function, it also moderately improves the temperature-dependent density of liquid water, and most strikingly, maintains close agreement with the geometries and binding energies of small water clusters. This illustrates the improved robustness of the QMFF potential to adapt to different environments due mainly to its explicit inclusion of quantum many-body response. Car–Parrinello simulations of liquid water have been performed with various DFT functionals in a plane-wave AO basis [213–215], and all of these studies inherently include explicit quantum many-body response; however, their O–O radial distribution functions are substantially worse than TIP4PEw. Specifically, Car–Parrinello DFT simulations produce radial distribution functions that are too structured [213–215]. This should not be surprising because TIP4PEw has been specifically parameterized to reproduce bulk properties of liquid water, whereas the DFT methods were not. QMFFs achieve high accuracy by parametrically tuning the intermolecular interactions; therefore, a QMFF may become more accurate than its underlying QM method for those properties that it is parameterized to reproduce.

### 6.2. Small molecule crystals

As an example of the solid state, QMFFs have been demonstrated in simulations of small-molecule crystals. Small molecule crystals are of pharmaceutical interest, as different crystalline polymorphs can result in different pharmacokinetic properties [21], and further are useful to study as fundamental benchmark systems that have experimentally well-characterized atomic structure and fluctuations. Figure 1 illustrates the average structure and atomic fluctuations for a crystal supercell of *n,n*-dimethylglycine containing 18 independent replicas of the fundamental unit cell (16 DMG molecules per unit cell, 288 DMG molecules in the supercell). Simulations were performed in the NVT ensemble with a 1 fs integration time step carried out to 220 ps at 275 K and using the mDC QMFF model with DFTB2 intra-fragment Hamiltonian, as described in detail elsewhere [60]. The average asymmetric unit structure from the crystal simulation had a root-mean-square (rms) displacement of 0.096 Å with the crystal structure, and the calculated fluctuations were highly correlated with the experimental Debye–Waller factors (B-factors). The mDC method with multipolar Ewald achieved a simulation rate of 160ps

per day using 16 cores (Intel Xeon E5-2670 @ 2.60 GHz), which is impressive for a large 4608-particle fully quantum mechanical simulation.

## 7. Challenges for the future

QMFFs offer a host of potential advantages—and disadvantages—relative to potential competing methods for condensed phase simulations. In comparison with LSQMs, for example, QMFFs are considerably faster using the same base level of theory, have no significant break-even point, and offer parametric freedom that can be tuned to obtain potentially more accurate condensed phase properties. However, the fragmentation approach that enables QMFFs to have greater computational speed-up also poses challenges in modeling reactive chemical events that involve bond formation and cleavage to occur between fragments. Thus, a direction of future work will involve creation of adaptive fragmentation approaches that allow general reactivity between fragments to occur smoothly, while maintaining a high level of computational efficiency.

In comparison with empirical force fields, including polarizable force fields, QMFFs have fewer parameters, particularly for bonded intrafragment interactions, and thus robust models can potentially be developed with greater ease. This advantage is useful for unconventional molecular systems (such as encountered in drug design) where empirical force field parameters may not be well tested, or in some cases may not even be available. Nonetheless, parameterization against condensed phase properties will still require considerable effort, particularly because the software infrastructure for performing simulations with QMFFs is still in its infancy. Along these same lines, extension of QMFFs to polymers that are covalently linked will require additional parameterization effort to ensure correct conformational distributions around rotatable bonds connecting fragments. Hence, a future area for research will involve developing robust automated methods for parameterizing and testing QMFFs for non-bonded interactions of arbitrary molecules, and general methods for determination of empirical terms required to connect fragments in polymer systems, and particularly for biopolymers.

While QMFFs offer tremendous speed advantages relative to fully QM methods, including LSQMs, they remain computationally inefficient relative to traditional force fields. Hence, an important area of future work will be to improve the performance of QMFFs through new methodology and software development, including optimization for acceleration on specialized hardware such as graphical processing units. Much effort has been made to develop GPU-accelerated software for molecular dynamics and quantum chemistry calculations. The latter currently poses special challenges due to the typically required limited precision of the GPUs, however the field remains dynamic, and this issue may become less of a restriction in the future with improvement of double precision performance.

Finally, the application scope of QMFFs within a multi-scale modeling paradigm is yet to be ascertained. Ultimately, QMFFs likely will see their greatest application within a

multiscale modeling framework, as opposed to as a substitution for conventional quantum methods and force fields. Potential directions for future work involve exploring combined QM, QMFF and MM models to optimize accuracy and performance for specific applications. As a further example, QMFFs could prove advantageous in refinement of MM free energy perturbation calculations when used as MM→QMFF endpoint corrections, as has been the recent focus of QM/MM free energy calculations [216–218]. Potential applications that likely will benefit from this approach include: (1) simulations of molecular crystals, (2) calculation of IR, Raman and NMR chemical shifts, and (3) docking of non-covalent and covalent ligands and drugs, (4) free energy calculations of metal-ion coupled  $pK_a$  shifts and redox potentials, and (5) combined QM/QMFF simulations of biocatalysis.

## Acknowledgments

The authors are grateful for financial support provided by the National Institutes of Health (No. GM107485). Computational resources were provided by the National Institutes of Health under Grant No. S10OD012346 and by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant No. OCI-1053575 (Project No. TG-MCB110101).

## References

- [1] Brooks B R *et al* 2009 *J. Comput. Chem.* **30** 1545–614
- [2] Halgren T A and Damm W 2001 *Curr. Opin. Struct. Biol.* **11** 236–42
- [3] Shi Y, Ren P, Schnieders M and Piquemal J-P 2015 *Polarizable Force Fields for Biomolecular Modeling (Reviews in Computational Chemistry vol 28)* (New York: Wiley) ch 2, pp 51–86
- [4] Lemkul J A, Huang J, Roux B and MacKerell A D 2016 *Chem. Rev.* **116** 4983–5013
- [5] Head-Gordon M 1996 *J. Phys. Chem.* **100** 13213–25
- [6] Friesner R A and Victor Guallar 2005 *Annu. Rev. Phys. Chem.* **56** 389–427
- [7] Cohen A J, Mori-Sánchez P and Yang W 2012 *Chem. Rev.* **112** 289–320
- [8] Jordan D K and Mazziotti D A 2005 *J. Chem. Phys.* **122** 084114
- [9] Németh K and Scuseria G E 2000 *J. Chem. Phys.* **113** 6035–41
- [10] Andermatt S, Cha J, Schiffmann F and VandeVondele J 2016 *J. Chem. Theory Comput.* **12** 3214–27
- [11] Weigand F and Kattannek M 2009 *J. Chem. Phys.* **130** 164106
- [12] Grimme S, Brandenburg J G, Bannwarth C and Hansen A 2015 *J. Chem. Phys.* **143** 054107–26
- [13] Dewar M J and Thiel W 1977 *J. Am. Chem. Soc.* **99** 4899–907
- [14] Elstner M, Porezag D, Jungnickel G J, Haugk M, Frauenheim T, Suhai S and Seifert G 1998 *Phys. Rev. B* **58** 7260–8
- [15] Gaus M, Goetz A and Elstner M 2013 *J. Chem. Theory Comput.* **9** 338–54
- [16] Gaus M, Lu X, Elstner M and Cui Q 2014 *J. Chem. Theory Comput.* **10** 1518–37
- [17] Giese T J and York D M 2008 *J. Comput. Chem.* **29** 1895–904
- [18] Watson M A, Salek P, Macak P and Helgaker T 2004 *J. Chem. Phys.* **121** 2915–31
- [19] Yang W and Pérez-Jordá J M 1998 *Encyclopedia of Computational Chemistry* ed P R von Schleyer (New York: Wiley) pp 1496–513
- [20] Goedecker S and Scuseria G E 2003 *IEEE Comput. Sci. Eng.* **5** 14–21
- [21] Giese T J, Huang M, Chen H and York D M 2014 *Acc. Chem. Res.* **47** 2812–20
- [22] Gao J, Truhlar D G, Wang Y, Mazack M J, Löffler P, Provorse M R and Rehak P 2014 *Acc. Chem. Res.* **47** 2837–45
- [23] Gao J 1996 *Acc. Chem. Res.* **29** 298–305
- [24] Gao J 1997 *J. Phys. Chem.* **101** 657–63
- [25] Ewig C S *et al* 2001 *J. Comput. Chem.* **22** 1782–800
- [26] Grimme S 2014 *J. Chem. Theory Comput.* **10** 4497–514
- [27] Sun L and Deng W-Q 2017 *WIREs Comput. Mol. Sci.* **7** e1282
- [28] Bowler D R and Gillan M J 2002 *Chem. Phys. Lett.* **355** 306–10
- [29] Knizia G and Chan G 2012 *Phys. Rev. Lett.* **109** 186404
- [30] Wouters S, Jiménez-Hoyos C A, Sun Q and Chan G K-L 2016 *J. Chem. Theory Comput.* **12** 2706–19
- [31] Cheng J, Libisch F, Yu K, Dieterich J M and Carter E A 2017 *J. Chem. Theory Comput.* **13** 1067–80
- [32] Cheng J, Yu K, Libisch F, Dieterich J M and Carter E A 2017 *J. Chem. Theory Comput.* **13** 1081–93
- [33] Kevorkyants R, Eshuis H and Pavanello M 2014 *J. Chem. Phys.* **141** 044127–37
- [34] Genova A and Pavanello M 2015 *J. Phys.: Condens. Matter* **27** 495501–7
- [35] Kristhal A, Sinha D, Genova A and Pavanello M 2015 *J. Phys.: Condens. Matter* **27** 183202–28
- [36] Thiel W and Voityuk A A 1992 *Int. J. Quantum Chem.* **44** 807–29
- [37] Dewar M J S, Zoebisch E, Healy E F and Stewart J J P 1985 *J. Am. Chem. Soc.* **107** 3902–9
- [38] Frauenheim T, Seifert G, Hajnal Z, Jungnickel G, Porezag D, Suhai S and Scholz R 2000 *Phys. Status Solidi b* **217** 41–62
- [39] Gaus M, Cui Q and Elstner M 2011 *J. Chem. Theory Comput.* **7** 931–48
- [40] Huang M, Giese T J and York D M 2015 *J. Comput. Chem.* **36** 1370–89
- [41] Otte N, Scholten M and Thiel W 2007 *J. Phys. Chem. A* **111** 5751–5
- [42] Huang M, Giese T J, Lee T-S and York D M 2014 *J. Chem. Theory Comput.* **10** 1538–45
- [43] Weber W and Thiel W 2000 *Theor. Chem. Acc.* **103** 495–506
- [44] Tuttle T and Thiel W 2008 *Phys. Chem. Chem. Phys.* **10** 2159–66
- [45] Korth M and Thiel W 2011 *J. Chem. Theory Comput.* **7** 2929–36
- [46] Kulik H J, Seelam N, Mar B D and Martínez T J 2016 *J. Phys. Chem. A* **120** 5939–49
- [47] Giese T J, Sherer E C, Cramer C J and York D M 2005 *J. Chem. Theory Comput.* **1** 1275–85
- [48] Nam K, Cui Q, Gao J and York D M 2007 *J. Chem. Theory Comput.* **3** 486–504
- [49] Cai Z-L, Lopez P, Reimers J R, Cui Q and Elstner M 2007 *J. Phys. Chem. A* **111** 5743–50
- [50] Stewart J J P 1989 *J. Comput. Chem.* **10** 221–64
- [51] Stewart J J P 2007 *J. Mol. Model.* **13** 1173–213
- [52] Zhang P, Fiedler L, Leverentz H R, Truhlar D G and Gao J 2011 *J. Chem. Theory Comput.* **7** 857–67
- [53] Mazack M J M and Gao J 2014 *J. Chem. Phys.* **140** 204501
- [54] Giese T J and York D M 2005 *J. Chem. Phys.* **123** 164108
- [55] Giese T J and York D M 2012 *Theor. Chem. Acc.* **131** 1145
- [56] Kaminski S, Giese T J, Gaus M, York D M and Elstner M 2012 *J. Phys. Chem. A* **116** 9131–41
- [57] Christensen A S, Elstner M and Cui Q 2015 *J. Chem. Phys.* **143** 084123

- [58] Domínguez A, Niehaus T A and Frauenheim T 2015 *J. Phys. Chem. A* **119** 3535–44
- [59] Giese T J, Chen H, Huang M and York D M 2014 *J. Chem. Theory Comput.* **10** 1086–98
- [60] Giese T J, Panteva M T, Chen H and York D M 2015 *J. Chem. Theory Comput.* **11** 451–61
- [61] Giese T J and York D M 2016 *A Modified Divide-and-Conquer Linear-Scaling Quantum Force Field With Multipolar Charge Densities* ed Q Cui *et al* (Singapore: Pan Stanford) ch 1
- [62] Kuechler E R, Giese T J and York D M 2016 *J. Chem. Phys.* **144** 164115
- [63] Giese T J and York D M 2011 *J. Chem. Phys.* **134** 194103
- [64] Kullgren J, Wolf M J, Hermansson K, Köhler C, Aradi B, Frauenheim T and Broqvist P 2017 *J. Phys. Chem.* **121** 4593–607
- [65] Hellström M, Jorner K, Bryngelsson M, Huber S E, Kullgren J, Frauenheim T and Broqvist P 2013 *Phys. Chem. C* **117** 17004–15
- [66] Liu X, Wahiduzzaman M, Oliveira A F, Heine T and Salahub D R 2016 *Theor. Chem. Acc.* **135** 168
- [67] Shi H, Koskinen P and Ramasubramaniam A 2017 *J. Phys. Chem. A* **121** 2497–502
- [68] Jahangiri S, Dolgonos G, Frauenheim T and Peslherbe G H 2013 *J. Chem. Theory Comput.* **9** 3321–32
- [69] Witek H A, Chou C-P, Mazur G Y, Irle S, Aradi B, Frauenheim T and Morokuma K 2016 *J. Chin. Chem. Soc.* **63** 57–68
- [70] Kaminski S, Gaus M and Elstner M 2012 *J. Phys. Chem. A* **116** 11927–37
- [71] Cabral do Couto P, Hollas D and Slavíček P 2015 *J. Chem. Theory Comput.* **11** 3234–44
- [72] Wong B M and Hsieh T H 2010 *J. Chem. Theory Comput.* **6** 3704–12
- [73] Modrzejewski M, Chałasiński G and Szczyński M M 2014 *J. Chem. Theory Comput.* **10** 4297–306
- [74] Smith J M, Alahmadi Y J and Rowley C N 2013 *J. Chem. Theory Comput.* **9** 4860–5
- [75] Seth M, Ziegler T, Steinmetz M and Grimme S 2013 *J. Chem. Theory Comput.* **9** 2286–99
- [76] Grimme S, Hansen A, Brandenburg J G and Bannwarth C 2016 *Chem. Rev.* **116** 5105–54
- [77] Grimme S, Antony J, Ehrlich S and Krieg H 2010 *J. Chem. Phys.* **132** 154104–23
- [78] Brandenburg J G and Grimme S 2014 *J. Phys. Chem. Lett.* **5** 1785–9
- [79] Brandenburg J G and Grimme S 2016 *Acta Cryst. B* **72** 502–13
- [80] Brandenburg J G and Grimme S 2014 *Top. Curr. Chem.* **345** 1–23
- [81] Giese T J *et al* 2013 *J. Chem. Theory Comput.* **9** 1417–27
- [82] Gordon M S, Fedorov D G, Pruitt S P and Slipchenko L V 2012 *Chem. Rev.* **112** 632–72
- [83] Richard R M and Herbert J M 2012 *J. Chem. Phys.* **137** 064113
- [84] Liu J and Herbert J M 2016 *J. Chem. Theory Comput.* **12** 572–84
- [85] He X, Zhu T, Wang X, Liu J and Zhang J 2014 *Acc. Chem. Res.* **47** 2748–57
- [86] Ramabhadran R O and Raghavachari K 2014 *Acc. Chem. Res.* **47** 3596–604
- [87] Collins M A and Bettens R P A 2015 *Chem. Rev.* **115** 5607–42
- [88] Raghavachari K and Saha A 2015 *Chem. Rev.* **115** 5643–77
- [89] Akimov A V and Prezhdo O V 2015 *Chem. Rev.* **115** 5797–890
- [90] Gordon M S, Mullin J M, Pruitt S R, Roskop L B, Slipchenko L V and Boatz J A 2009 *J. Phys. Chem. B* **113** 9646–63
- [91] Fedorov D G and Kitaura K 2007 *J. Phys. Chem. A* **111** 6904–14
- [92] Morokuma K 2002 *Phil. Trans. R. Soc. Lond. A* **360** 1149–64
- [93] Lee C and Yang W 1992 *J. Chem. Phys.* **96** 2408–11
- [94] Lee T-S, York D M and Yang W 1996 *J. Chem. Phys.* **105** 2744–50
- [95] Lee T-S and Yang W 1998 *Int. J. Quantum Chem.* **69** 397–404
- [96] Gogonea V, Westerhoff L M and Merz K M Jr 2000 *J. Chem. Phys.* **113** 5604–13
- [97] Hu H, Lu Z, Elstner M and Hermans J 2007 *J. Phys. Chem. A* **111** 5685–91
- [98] Song G-L, Li Z H, Liu Z-P, Cao X-M, Wang W, Fan K-N, Xie Y and Schaefer H F III 2008 *J. Chem. Theory Comput.* **4** 2049–56
- [99] Alizadegan R, Hsia K J and Martinez T J 2010 *J. Chem. Phys.* **132** 034101
- [100] He X and Merz K M Jr 2010 *J. Chem. Theory Comput.* **6** 405–11
- [101] Kobayashi M, Kunisada T, Akama T, Sakura D and Nakai H 2011 *J. Chem. Phys.* **134** 034105
- [102] Monard G, Bernal-Uruchurtu M I, van der Vaart A, Merz K M Jr and Ruiz-López M F 2005 *J. Phys. Chem. A* **109** 3425–32
- [103] Jensen J H, Li H, Robertson A D and Molina P A 2005 *J. Phys. Chem. A* **109** 6634–43
- [104] Wang B, Brothers E N, van der Vaart A and Merz K M Jr 2004 *J. Chem. Phys.* **120** 11392–400
- [105] Wang B, Raha K and Merz K M Jr 2004 *J. Am. Chem. Soc.* **126** 11430–1
- [106] Nagata T, Brorsen K, Fedorov D G, Kitaura K and Gordon M S 2011 *J. Chem. Phys.* **134** 124115
- [107] Fedorov D G, Nagata T and Kitaura K 2012 *Phys. Chem. Chem. Phys.* **14** 7562–77
- [108] Mayhall N J and Raghavachari K 2011 *J. Chem. Theory Comput.* **7** 1336–43
- [109] Brorsen K R, Minezawa N, Xu F, Windus T L and Gordon M S 2012 *J. Chem. Theory Comput.* **8** 5008–12
- [110] Nakata H, Fedorov D G, Nagata T, Kitaura K and Nakamura S 2015 *J. Chem. Theory Comput.* **11** 3053–64
- [111] Gao J and Wang Y 2012 *J. Chem. Phys.* **136** 071101
- [112] Slipchenko L V 2016 *Many-Body Effects, Electrostatic in Biomolecules* ed Q Cui *et al* (Singapore: Pan Stanford) ch 5, pp 147–90
- [113] Kumar Gurunathan P, Ghosh D, Kosenkov D, Kaliman I, Shao Y, Krylov A I and Slipchenko L V 2016 *J. Phys. Chem. B* **120** 6562–74
- [114] Gordon M S, Slipchenko L, Li H and Jensen J H 2007 *Annual Reports in Computational Chemistry* (Amsterdam: Elsevier) pp 177–93
- [115] Chaudret R, Ulmer S, van Severen M-C, Gresh N, Parisel O, Cisneros G A, Darden T A and Piquemal J-P 2008 *Theory, Applications of Computational Chemistry* ed D-Q Wei and X-J Wang (New York: American Institute of Physics) pp 185–92
- [116] Piquemal J and Chevreau H 2007 *J. Chem. Theory Comput.* **3** 824–37
- [117] Gresh N, Cisneros G A, Darden T A and Piquemal J-P 2007 *J. Chem. Theory Comput.* **3** 1960–86
- [118] Piquemal J-P, Gresh N and Giessner-Prettre C 2003 *J. Phys. Chem. A* **107** 10353–9
- [119] Gresh N, Claverie P and Pullman A 1984 *Theor. Chem. Acc.* **66** 1–20
- [120] Piquemal J-P and Cisneros G A 2016 Chapter: Status of the gaussian electrostatic model, a density-based polarizable force field *Many-Body Effects and Electrostatics* (Singapore: Pan Stanford) pp 269–99
- [121] Piquemal J, Cisneros G, Gresh N and Darden T A 2006 *J. Chem. Phys.* **124** 104101
- [122] Duke R E, Starovoytov O N, Piquemal J-P and Cisneros G A 2014 *J. Chem. Theory Comput.* **10** 1361–5

- [123] Cisneros G A, Piquemal J-P and Darden T A 2006 *J. Phys. Chem. B* **110** 13682–4
- [124] Tang K T and Toennies J P 1997 *J. Chem. Phys.* **66** 1496–506
- [125] Murrell J N and Teixeira-Dias J J C 1970 *Mol. Phys.* **19** 521–31
- [126] Wheatley R J and Price S L 1990 *Mol. Phys.* **69** 507–33
- [127] Xie W, Orozco M, Truhlar D G and Gao J 2009 *J. Chem. Theory Comput.* **5** 459–67
- [128] Xie W and Gao J 2007 *J. Chem. Theory Comput.* **3** 1890–900
- [129] Song L, Han J, Lin Y, Xie W and Gao J 2009 *J. Phys. Chem. A* **113** 11656–64
- [130] Han J, Truhlar D G and Gao J 2012 *Theor. Chem. Acc.* **131** 1161
- [131] Wang Y, Sosa C P, Cembran A, Truhlar D and Gao J 2012 *J. Phys. Chem. B* **116** 6781–8
- [132] Han J, Mazack M J M, Zhang P, Truhlar D G and Gao J 2013 *J. Chem. Phys.* **139** 054503
- [133] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 *J. Chem. Phys.* **79** 926–35
- [134] Pu J, Jiali Gao P, Donald G and Truhlar P 2005 *Chem. Phys. Chem.* **6** 1853–65
- [135] Pu J, Gao J and Truhlar D G 2004 *J. Phys. Chem. A* **108** 632–50
- [136] Garcia-Viloca M and Gao J 2004 *Theor. Chem. Acc.* **111** 280–6
- [137] Pu J, Gao J and Truhlar D G 2004 *J. Phys. Chem. A* **108** 5454–63
- [138] Amara P, Field M J, Alhambra C and Gao J 2000 *Theor. Chem. Acc.* **104** 336–43
- [139] Gao J, Amara P, Alhambra C and Field M 1998 *J. Phys. Chem. A* **102** 4714–21
- [140] Walker R C, Crowley M F and Case D A 2008 *J. Comput. Chem.* **29** 1019–31
- [141] Lin H and Truhlar D 2005 *J. Phys. Chem. A* **109** 3991–4004
- [142] Amara P and Field M J 2003 *Theor. Chem. Acc.* **109** 43–52
- [143] DiLabio G A, Hurly M M and Christiansen P A 2002 *J. Chem. Phys.* **116** 9578–84
- [144] Wang X, Liu J, Zhang J Z H and He X 2013 *J. Phys. Chem. A* **117** 7149–61
- [145] Gao A M, Zhang D W, Zhang J Z and Zhang Y 2004 *Chem. Phys. Lett.* **394** 293–7
- [146] Liu J, Zhu T, Wang X, He X and Zhang J Z H 2015 *J. Chem. Theory Comput.* **11** 5897–905
- [147] Cembran A, Bao P, Wang Y, Truhlar D G and Gao J 2010 *J. Chem. Theory Comput.* **6** 2469–76
- [148] Senftle T P *et al* 2016 *NPJ Comput. Mater.* **2** 15011–24
- [149] van Duin A C, Dasgupta S and Goddard W A III 2001 *J. Phys. Chem. A* **105** 9396–409
- [150] Nielson K D, van Duin A C, Oxgaard J, Deng W-Q and Goddard W A III 2005 *J. Phys. Chem. A* **109** 493–9
- [151] Rice B M, Larentzos J P, Byrd E F C and Weingarten N S 2015 *J. Chem. Theory Comput.* **11** 392–405
- [152] Müller J and Hartke B 2016 *J. Chem. Theory Comput.* **12** 3913–25
- [153] Larentzos J P and Rice B M 2017 *J. Phys. Chem. A* **121** 2001–13
- [154] Han Y, Jiang D, Zhang J, Li W and Gu J 2016 *Front. Chem. Sci. Eng.* **10** 16–38
- [155] Ewald P P 1921 *Ann. Phys.* **64** 253–68
- [156] Darden T, York D and Pedersen L 1993 *J. Chem. Phys.* **98** 10089–92
- [157] Essmann U, Perera L, Berkowitz M L, Darden T, Hsing L and Pedersen L G 1995 *J. Chem. Phys.* **103** 8577–93
- [158] Cerutti D S, Duke R E, Darden T A and Lybrand T P 2009 *J. Chem. Theory Comput.* **5** 2322–38
- [159] Cerutti D S and Case D A 2010 *J. Chem. Theory Comput.* **6** 443–58
- [160] York D M, Wlodawer A, Pedersen L G and Darden T 1994 *Proc. Natl Acad. Sci. USA* **91** 8715–8
- [161] Cheatham T E III, Miller J L, Fox T, Darden T A and Kollman P A 1995 *J. Am. Chem. Soc.* **117** 4193–4
- [162] Cheatham T E III and Kollman P A 1996 *J. Mol. Biol.* **259** 434–44
- [163] Rick S W 2004 *J. Chem. Phys.* **120** 6085–93
- [164] Horn H W, Swope W C, Pitera J W, Madura J D, Dick T J, Hura G L and Head-Gordon T 2004 *J. Chem. Phys.* **120** 9665–78
- [165] van der Spoel D and van Maaren P J 2006 *J. Chem. Theory Comput.* **2** 1–11
- [166] Lee F S and Warshel A 1992 *J. Chem. Phys.* **97** 3100–7
- [167] Giese T J, Panteva M T, Chen H and York D M 2015 *J. Chem. Theory Comput.* **11** 436–50
- [168] Simmonett A C, Pickard F C IV, Schaefer H F III and Brooks B R 2014 *J. Chem. Phys.* **140** 184101
- [169] Narth C, Lagardère L, Polack É, Gresh N, Wang Q, Bell D R, Rackers J A, Ponder J W, Ren P and Piquemal J-P 2016 *J. Comput. Chem.* **37** 494–506
- [170] Nam K, Gao J and York D M 2005 *J. Chem. Theory Comput.* **1** 2–13
- [171] Zhang P, Truhlar D G and Gao J 2012 *Phys. Chem. Chem. Phys.* **14** 7821–9
- [172] Giese T J and York D M 2016 *J. Chem. Theory Comput.* **12** 2611–32
- [173] Holden Z C, Richard R M and Herbert J M 2013 *J. Chem. Phys.* **139** 244108
- [174] Holden Z C, Richard R M and Herbert J M 2015 *J. Chem. Phys.* **142** 059901
- [175] Kratz E G, Duke R E and Cisneros G A 2016 *Theor. Chem. Acc.* **135** 166
- [176] Fang D, Duke R E and Cisneros G A 2015 *J. Chem. Phys.* **143** 044103
- [177] Vasilevskaya T and Thiel W 2016 *J. Chem. Theory Comput.*
- [178] Riccardi D, Schaefer P and Cui Q 2005 *J. Phys. Chem. B* **109** 17715–33
- [179] Gao J and Alhambra C 1997 *J. Chem. Phys.* **107** 1212–7
- [180] Breneman C M and Wiberg K B 1990 *J. Comput. Chem.* **11** 361–73
- [181] Jacobson L D and Herbert J M 2011 *J. Chem. Phys.* **134** 094118
- [182] Herbert J M, Jacobson L D, Laoa K U and Rohrdanza M A 2012 *Phys. Chem. Chem. Phys.* **14** 7679–99
- [183] Hu H, Lu Z and Yang W 2007 *J. Theory Comput.* **3** 390–406
- [184] Hu H, Lu Z, Parks J M, Burger S K and Yang W 2008 *J. Chem. Phys.* **128** 034105
- [185] Zeng X, Hu H, Hu X, Cohen A J and Yang W and Yang W 2008 *J. Chem. Phys.* **128** 124510
- [186] Alper H and Levy R M 1993 *J. Chem. Phys.* **99**
- [187] Adams D J and Adams E M 1979 *J. Mol. Phys.* **38** 387–400
- [188] Woodcock H L III, Hodoscek M, Gilbert A T, Gill P M, Schaefer H F III and Brooks B R 2007 *Chem.* **28** 1485–502
- [189] Cui Q and Karplus M 2001 *J. Am. Chem. Soc.* **123** 2284–90
- [190] Cisneros G A, Liu H, Zhang Y and Yang W 2003 *J. Am. Chem. Soc.* **125** 10384–93
- [191] Liu H, Zhang Y and Yang W 2000 *J. Chem. Soc.* **122** 6560–70
- [192] Zhang S, Ganguly A, Goyal P, Bingaman J L, Bevilacqua P C and Hammes-Schiffer S 2015 *J. Am. Chem. Soc.* **137** 784–98
- [193] Thaplyal P, Ganguly A, Hammes-Schiffer S and Bevilacqua P C 2015 *Biochemistry* **54** 2160–75
- [194] Ganguly A, Thaplyal P, Rosta E, Bevilacqua P C and Hammes-Schiffer S 2014 *J. Am. Chem. Soc.* **136** 1483–96
- [195] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C 2006 *Proteins* **65** 712–725

- [196] Li J, Haycraft C and Iyengar S S 2016 *J. Chem. Theory Comput.* **12** 2493–508
- [197] Kuroki N and Mori H 2016 *Chem. Lett.* **45** 1009–11
- [198] Pruitt S R, Nakata H, Nagata T, Mayes M, Alexeev Y, Fletcher G, Fedorov D G, Kitaura K and Gordon M S 2016 *J. Chem. Theory Comput.* **12** 1423–35
- [199] Mazanetz M P, Ichihara O, Law R J and Whittaker M 2011 *J. Cheminform.* **3** 2–16
- [200] Lange A W and Voth G A 2013 *J. Chem. Theory Comput.* **9** 4018–25
- [201] Zeng X, Hu X and Yang W 2012 *Chem. Theory Comput.* **8** 4960–7
- [202] Isegawa M, Gao J and Truhlar D G 2011 *J. Chem. Phys.* **135** 084107
- [203] Nishizawa H, Nishimura Y, Kobayashi M, Irle S and Nakai H 2016 *J. Comput. Chem.* **37** 1983–92
- [204] Nakata H, Fedorov D G, Yokojima S, Kitaura K and Nakamura S 2014 *J. Chem. Theory Comput.* **10** 3689–98
- [205] Jovan Jose K V, Beckett D and Raghavachari K 2015 *J. Chem. Theory Comput.* **11** 4238–47
- [206] Hartman J D and Beran G J O 2014 *J. Chem. Theory Comput.* **10** 4862–72
- [207] Hartman J D, Kudla R A, Day G M, Mueller L J and Beran G J O 2016 *Phys. Chem. Chem. Phys.* **18** 21686–709
- [208] Gao Q, Yokojima S, Fedorov D G, Kitaura K, Sakurai M and Nakamura S 2010 *J. Chem. Theory Comput.* **6** 1428–44
- [209] Steinmann C, Bratholm L A, Olsen J M H and Kongsted J 2017 *J. Chem. Theory Comput.* **13** 525–36
- [210] Jin X, Zhang J Z H and He X 2017 *J. Phys. Chem. A* **121** 2503–14
- [211] Liu J, Qi L-W, Zhang J Z H and He X 2017 *J. Chem. Theory Comput.* **13** 2021–34
- [212] Akimov A V 2016 *J. Chem. Theory Comput.* **12** 5719–36
- [213] Khaliullin R Z and Kühne T D 2013 *Phys. Chem. Chem. Phys.* **15** 15746–66
- [214] Guardia E, Skarmoutsos I and Masia M 2015 *J. Phys. Chem. B* **119** 8926–38
- [215] Todorova T, Seitsonen A P, Hutter J, Kuo I-F W and Mundy C J 2006 *J. Phys. Chem. B* **110** 3685–91
- [216] Sampson C, Fox T, Tautermann C S, Woods C and Skylaris C-K 2015 *J. Phys. Chem. B* **119** 7030–40
- [217] König G, Hudson P S, Boresch S and Woodcock H L 2014 *J. Chem. Theory Comput.* **10** 1406–19
- [218] König G *et al* 2016 *J. Comput. Aided Mol.* **30** 989–1006
- [219] Temelso B, Archer K A and Shields G C 2011 *J. Phys. Chem. A* **115** 12034–46
- [220] Kell G S 1975 *J. Chem. Eng. Data* **20** 97–105
- [221] Feistel R and Wagner W 2006 *J. Phys. Chem. Ref. Data* **35** 1021–47
- [222] Skinner L B, Huang C, Pettersson L G M, Nilsson A and Benmore C J 2013 *J. Chem. Phys.* **138** 074506