Postprint of: Albaugh A., Boateng H., Bradshaw R., Demerdash O., Dziedzic J., Mao Y., Margul D., Swails J., Zeng Q., Case D., Eastman P., Essex J., Head-Gordon M., Pande V., Ponder J., Shao Y., Skylaris C., Todorov I., Tuckerman M., Head-Gordon T., Advanced Potential Energy Surfaces for Molecular Simulation, JOURNAL OF PHYSICAL CHEMISTRY B, Vol. 120, Iss. 37 (2016), pp.9811-9832

Advanced Potential Energy Surfaces for Molecular Simulation

Alex Albaugh[3†], Henry A. Boateng[9], Richard T. Bradshaw[1†], Omar N. Demerdash[2†], Jacek Dziedzic[1,5†], Yuezhi Mao[2†], Daniel T. Margul[6†], Jason Swails[10†], Qiao Zeng[8†], David A. Case[10], Peter Eastman[11], Jonathan W. Essex[1], Martin Head-Gordon[2], Vijay S. Pande[11], Jay W. Ponder[12], Yihan Shao[13], Chris-Kriton Skylaris[1], Ilian T. Todorov[14], Mark E. Tuckerman[6,7,15], Teresa Head-Gordon[2,3,4*]

[1]*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK.*
[2]*Department of Chemistry,* [3]*Chemical and Biomolecular Engineering, and* [4]*Bioengineering, University of California, Berkeley, CA 94720*

[5]*Faculty of Applied Physics and Mathematics, Gdansk University of Technology, Poland*
[6]*Department of Chemistry and* [7]*Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA*

[8]*Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892*
[9]*Department of Mathematics, Bates College, 2 Andrews Road, Lewiston, ME 04240*
[10]*Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854-8066, United States*
[11]*Department of Chemistry, Stanford University, Stanford, CA 94305*
[12]*Department of Chemistry, Washington University in St. Louis, St. Louis, Missouri, 63130*

[13]*Q-Chem Inc., 6601 Owens Drive, Suite 105, Pleasanton, California 94588*
[14]*STFC Daresbury Laboratory, Keckwick Lane, Daresbury, Warrington WA4 4AD, UK*
[15]*NYU-ECNU, Center for Computational Chemistry at NYU, Shanghai, Shanghai 200062, China*

ABSTRACT

Advanced potential energy surfaces are defined as theoretical models that explicitly include many-body effects that transcend the standard fixed-charge, pairwise-additive paradigm typically used in molecular simulation. However, several factors relating to their software implementation have precluded their widespread use in condensed-phase simulations: the computational cost of the theoretical models, a paucity of approximate models and algorithmic improvements that can ameliorate their cost, under-developed interfaces and limited dissemination in computational code bases that are widely used in the computational chemistry community, and software implementations that have not kept pace with modern high-performance computing (HPC) architectures, such as multicore CPUs and modern graphics processing units (GPUs). In this Feature article we review recent progress made in these areas, including well-defined polarization approximations and new multipole electrostatic formulations, novel methods for solving the mutual polarization equations and increasing the MD time step, combining linear scaling electronic structure methods with new QM/MM methods that account for mutual polarization between the two regions, and the greatly improved software deployment of these models and methods onto GPU and CPU hardware platforms. We have now approached an era where multipole-based polarizable force fields can be routinely used to obtain computational results comparable to state-of-the-art density functional theory while reaching sampling statistics th e when compared to that obtained from simpler fixed partial charge force fields.

*corresponding author: thg@berkeley.edu, (510) 666-2744; †authors contributed equally

1. INTRODUCTION

Over the last 50 years computational chemistry has advanced to be an equal partner with experiment in research areas ranging from lead optimization in drug discovery through to mechanistic insight into catalysts such as zeolites and enzymes. Historically, these successes have relied on the most tractable classical models for condensed phase simulation: the assumption of a pairwise additive fixed charge force field, whose functional form is nearly identical to that laid out by Lifson and Warshel in the late 1960's[1].

$$U = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} k_\phi \left(1 + cos(n\phi - \omega)\right) + \tag{1}$$

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \frac{q_i q_j}{r_{ij}}$$

Pairwise additive molecular models are widely available in a range of community codes such as Amber[2], Charmm[3], NAMD[4], and OpenMM[5], and their success is due to the continued improvements in optimization of their parameters through a pragmatic approach of comparison to quantum chemical calculations and empirical fitting to condensed phase properties. Although pairwise additive models may be an inadequate representation of the true many-body physics of the quantum mechanical energy surface, their popularity also stems from the fact that they better permit the high dimensional spatial or temporal averaging that is dictated by the laws of statistical mechanics, especially since the software that implements them is also well-optimized on modern day computer architectures.

Before addressing the inadequacy of pairwise additivity, the molecular simulation field required a refractory period to sort out other important aspects of the molecular simulation protocol to generate meaningful results and analysis. This included overcoming finite size system effects, the use of Ewald summation for long-ranged electrostatics[6-7], integrators for the equations of motion that are symplectic[8], and extended system methods that formally reach the correct limiting thermodynamic ensemble[9-10]. With the advent of greater computing power combined with these improved simulation protocols, it then became possible to diagnose when the pairwise additive potential energy and forces were breaking down. The failures of pairwise additivity are unambiguous when one considers the "asymmetric environment"[11] such as the heterogeneity at interfaces[12-13], calculation of electric fields in complex protein environments[14-15], hydration free energies of a large range of small molecules[16-17], or aggregation propensities of hydrophobic peptides[18-20]. From this accumulating experience it is becoming apparent that we are reaching a generational transition in how to model the underlying potential energy surface. Almost

all of the leading force field development teams are devising transferable many-body force fields[20-50], and advances are being made to include more intricate molecular interactions that account for charge penetration[38, 51-53] and charge transfer[54-59]. In order to gain the full advantage of these advanced classical potential energy surfaces, they should also be properly combined with explicit quantum mechanical treatments, to yield better predictions when bond making and bond breaking are important in the condensed phase.

However, the extension to better physics comes at a cost. First, advanced potential energy surfaces carry a larger computational overhead such that it becomes more difficult to realize statistical convergence of condensed-phase properties. Second, the more complicated functional forms are harder to parameterize and hence are more brittle in their application, and new advances are needed to overcome the current limitations of hand-tuning parameters. Third, advanced treatments of electron-electron interactions using QM/MM methods further increase computational expense and thus limit necessary sampling. Finally, the software implementation of advanced potential energy surfaces on current or emergent hardware platforms has posed several challenges that preclude their widespread adoption in the computational chemistry community.

These concurrent issues can be illustrated using the status of the classical polarization force field AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications)[27, 60-61] and its implementation in the TINKER package that we reviewed previously in 2010[62]. AMOEBA has the following functional form for the interactions among atoms

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{oop} + U_{torsion} + U_{VDW} + U_{elec}^{perm} + U_{elec}^{ind} \qquad (2)$$

where the first six terms describe the short-range valence interactions such as bond and angle deformations, bond-angle cross terms, a formal Wilson-Decius-Cross decomposition of angle bending into in-plane and out-of-plane components, and a "softer" buffered 14-7 van der Waals form. In order to provide a better description of short-ranged anisotropic interactions, atomic multipoles and polarizable dipoles replace the standard fixed partial charges in the last two terms in Eq. (2). The essence of the AMOEBA model can be effectively captured in the polarization equation for the induced dipole vector on polarizable site *i*

$$\mu_{i,\gamma}^{(d)} = \alpha_i \left(\sum_j^{13N} T_{ij,\gamma} M_j^{(d)} + \sum_k^{3N} T'_{ik,\gamma\delta} \mu_{k,\delta} \right) \qquad \gamma, \delta = x, y, z \qquad (3)$$

where $\vec{\mu}_i$ is the inducible dipole at atom site *i*, α_i is the isotropic polarizability of atom *i*, T_{ij} is the rank-two interaction tensor between atoms *i* and *j* containing derivatives of $1/r_{ij}$ prescribed by the permanent

multipole expansion, T_{ik}' is the corresponding interaction tensor just for inducible dipole-dipole interactions, and $M_j^{(d)}$ are the permanent multipole moments. The rank-two tensor T prescribes the interaction among the permanent multipole sites through quadrupole given in the rank-one tensor M (q, μ_x, μ_y, μ_z, Q_{xx}, Q_{xy}, Q_{xz}, Q_{yx}, Q_{yy}, Q_{yz}, Q_{zx}, Q_{zy}, Q_{zz}); the d superscript refers to scaling factors in the AMOEBA potential that prescribe which permanent multipole sites may give rise to polarization in another site. The fixed multipoles introduce significant computational overhead with respect to standard fixed charge models, and historically the additional cost of solving the linear set of polarization equations using a self-consistent field approach based on the conservative successive overrelaxation method[63] have contributed to slow adoption of AMOEBA by the molecular simulation community.

The TINKER codes from the Ponder group[64], the pmemd.amoeba code in Amber (primarily written by Bob Duke)[2], and ffx[65], have been the main community codes supporting the AMOEBA force field for the last two decades. On a single thread, such codes yield roughly 0.04 ns/day for the 161 amino acid cyclophilin A protein embedded in 6149 water molecules (this particular timing is for pmemd.amoeba on an Intel Xeon E5-2650 at 2.3 GHz, using a 1 fs time step). The OpenMM implementation of AMOEBA provides significant speed improvements, running this example at 1.7 ns/day on an NVIDIA GTX980 GPU (using mixed precision and again assuming a 1 fs time step). Leveraging a standard RESPA[66] multiple time step method can increase the time step to 2 fs and parallelization strategies via OpenMP or MPI calls speeds up this calculation by at most a factor of ~10, so that tens to hundreds of nanosecond simulations would require months of time even on high-end compute clusters. Still, relatively few studies using AMOEBA have been carried out on simulation time scales that are now considered routine for simpler, fixed-charge force fields. As an example of relative speed performance, the fixed charge Amber MD simulations for cyclophilin A runs at 486 ns/day on an NVIDIA GTX980 GPU, using hydrogen mass repartitioning[67], and SHAKE and SETTLE for constraining bonds to hydrogens, thereby permitting a 4 fs time step. Although this comparison combines differences in computational complexity with differences in hardware capability, it represents in a general way the practical speed tradeoffs facing users of these codes in early 2016.

It also illustrates an important, and hopefully as we show here, a transient observation: although the performance penalty for going from a simple fixed-charge force field to a more complex, polarizable potential like AMOEBA should in theory be one order of magnitude or less, these implementations yield performance ratios closer to ~140 (486/3.4) on GPUs and up to ~600 (486/0.8) on CPUs with MPI, and assuming a 2 fs time step as illustrated here for the cyclophilin A protein. The effective ratio between AMOEBA and a fixed charge force field can be even larger if one takes into account that the use of rigid

water models, which is very common with simpler force fields, and which permits even longer time steps. However, fixed geometry models are not philosophically in line with advanced force fields such as AMOEBA. Therefore, it is clear that further model innovation, new method development, and greatly improved software implementations are needed if more routine adoption of multipole-based polarizable force fields is to become a reality.

This Feature article presents a summary of recent results by a large consortium of researchers, which indicate that the state-of-the-art is changing rapidly, thus lowering the barrier to the more standard use of advanced force fields for large systems on long timescales. In Section 2 we introduce the range of new models and model physics centered around the AMOEBA force field. In section 2.1 we lay out a many-body expansion (MBE) formalism that provides well-defined approximations to complete mutual polarization, including the direct polarization model iAMOEBA[68] and the approximate mutual polarization model 3-AMOEBA[69], both of which yield acceptable accuracy and significant computational speed-ups over the parent AMOEBA potential. In sections 2.2 and 2.3 we describe two new QM/AMOEBA models that account for true mutual polarization across the QM/MM boundary as implemented in ONETEP[70] and Q-Chem[71]. In Section 3 we summarize new methodological advances that tackle important algorithmic bottlenecks for simulating advanced potential energy surfaces. In sections 3.1 we first consider the problem of more efficient sampling of AMOEBA through the formulation of a combined extended Lagrangian and self-consistent field solver for mutual polarization[72] and in Section 3.2 we describe a multiple time-step (MTS) integration algorithm that allows time steps of ~100 fs to be employed in molecular dynamics simulations of AMOEBA[73]. In Section 3.3 we describe improvements in Particle Mesh Ewald (PME) multipole electrostatics as implemented in DL_POLY[74] and in Section 3.4 we describe the progress made in more tractable linear scaling of density functional theory[75] that can be usefully combined with the QM/AMOEBA model described in sections 2.2 and 2.3. In Section 4 we describe multiple new GPU and CPU implementations of AMOEBA and other approximate AMOEBA models and demonstrate their use in a range of application areas. Section 4.1 presents modest improvements in the OpenMP version of AMOEBA in TINKER, and Section 4.2 describes a hybrid OpenMP/MPI implementation of the iAMOEBA, 3-AMOEBA, and AMOEBA models implemented in TINKER7 and TINKER-HP. In Section 4.3 we highlight the recent advances of fast GPU formulations of AMOEBA in OpenMM[5]. In the final Section 5 we present validation and some large molecular simulation results using AMOEBA and the improved code bases. Section 5.1 presents important validation studies of the AMOEBA potential including energy decomposition analysis of simple ion-water interactions compared against high quality density functional theory[76], and showing that simulated

proteins using the AMOEBA force field are structurally stable on long timescales of 100-500 ns. Given these validations, Section 5.2 describes the recent success of AMOEBA in SAMPL competitions for host-guest ligand predictions, and the ability to reproduce active site electrostatics previously quantified by much more expensive *ab initio* calculations on structures extracted from molecular dynamics simulations. We close in Section 6 with a final summary and conclusions.

SECTION 2. NEW MODELS OF ADVANCED POTENTIAL ENERGY SURFACES

2.1 A Hierarchy of Models based on the AMOEBA Force Field

The many-body expansion (MBE) to the total potential energy of an N-body potential[77-82]

$$U = U_1 + U_2 + U_3 + \cdots \tag{4a}$$

where

$$U_1 = \sum_{i=1}^{N} U(i), \quad U_2 = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} U(i,j) - U(i) - U(j), \tag{4b}$$

$$U_3 = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^{N} U(i,j,k) - U(i,j) - U(i,k) - U(j,k) + U(i) + U(j) + U(k)$$

provides a useful organizing principle for AMOEBA-based models and algorithms that alter the tradeoff between accuracy and computational speed, thereby allowing users to define a "sweet spot" for a given scientific application. The N-body AMOEBA polarizable force field can be evaluated as the set of atomic or molecular interactions in increasingly larger clusters starting from monomers, progressing to dimers, trimers, etc. When Eq. (4) is truncated at the level of trimers it defines direct polarization exactly (iAMOEBA) and mutual polarization approximately (3-AMOEBA), and in the limit of large N reduces to the complete AMOEBA polarization model. For the sake of clarification, a monomer may represent a small molecule such as an ion, a large water cluster, or a multi-atom fragment within a larger covalently bound structure like a biopolymer.

One of our primary goals in the development of iAMOEBA is to retain as much of the excellent AMOEBA model performance as possible, while reducing its computational cost for conditions where the number of degrees of freedom is large and when extensive statistical sampling is necessary. The iAMOEBA model is defined by a simple one step polarization scheme where the inducible dipoles respond only to the local field due to permanent atomic multipoles, *i.e.*

$$\mu_{i,\gamma}^{(d)} = \alpha_i \sum_{j}^{13N} T_{ij,\gamma} M_j^{(d)} \qquad \gamma, \delta = x, y, z \tag{5}$$

The question we addressed in previous work is whether the more limited direct polarization functional form in Eq. (5) can be reparameterized to recapture the missing mutual polarization. Using the automated parameterization tool ForceBalance [45, 83-84], we optimized the parameters of the iAMOEBA water model to recover the temperature dependence of the density and heat of vaporization, vibrational frequencies and other properties of gas phase water clusters[68]. Encouragingly, in studies of properties for which it was not explicitly parameterized such as the temperature dependence of the dielectric constant and the ordering of the ice phases[85], the iAMOEBA model proved to be a high quality water model, although it will require the wholesale reparameterization of other systems, including proteins, ions, and nucleic acids in order to broaden its applicability.

Nonetheless, there are good reasons to undertake the extensive parameterization work needed to extend iAMOEBA because the direct polarization form offers some significant advantages over full mutual polarization: (1) the iAMOEBA polarizable model is still responsive to the molecular environment at the direct polarization level, (2) iAMOEBA eliminates the expense of a tight SCF convergence for the induced dipoles or the poor accuracy associated with loose convergence of the dipoles (~10^{-2} D as was a typical convergence criterion in early simulations with the AMOEBA model), (3) iAMOEBA eliminates polarization catastrophes, for which full mutual polarization schemes require a damping function, and (4) there are no stability or accuracy issues in solving mutual polarization through an extended Langrangian formulation as used with many point dipole or Drude polarization models.

At the same time there is evidence that there are limitations to what can be captured by the direct polarization response. In particular, iAMOEBA energies and forces for gas phase water cluster are not as accurate compared to the full AMOEBA03 water model, the ratio of the dielectric constant of the liquid compared to that of ice is smaller than seen experimentally, and the second virial coefficients are too negative. What this suggests is that objective parameterization schemes like ForceBalance[83-84] have exhausted what is possible under the limited directed polarization functional form, and that mutual polarization is required to describe the wide-range of electric fields that differ substantially in the gas, liquid, and solid phases, and presumably are harbingers of problems at molecular interfaces between protein and water or at asymmetric environments such as the air-liquid surface. To that end, we have extended the ForceBalance approach to full mutual polarization, yielding the AMOEBA14 water model[86]. Overall, properties either remained as accurate as iAMOEBA, or they were significantly better. Examples of the latter are the interaction energy of gas-phase clusters ranging from dimers to 20-mers, the size-

corrected diffusion constant, and the second virial coefficients, all of which are in better agreement with experiment[86].

Although the demonstrably improved physics of a polarizable force field such as AMOEBA is a necessary pre-requisite for moving beyond the pairwise additive approximation, it is not sufficient if the computational expense of an energy and force calculation remains a huge bottleneck for AMOEBA. Therefore, it is prudent to approximate mutual polarization by means of a truncation of Eq. (4) at the level of the 3-body term; this defines the 3-AMOEBA model. The 3-AMOEBA approximation rests on the reduced computational cost of polarization for the individual fragments, as well as the fact that these individual subsystems may be calculated independently, allowing for trivial parallelization with little communication overhead. The fact that 2- and 3-body polarization energies decay rapidly with distance permits the use of inter-fragment distance cutoffs, reducing a potentially $O(N^3)$ computation to one that scales tractably as $O(N)$.

The implementation of 3-AMOEBA was tested for its ability to reproduce accurate condensed-phase energies as well as structural properties of water in the NVT and NPT ensembles, in which a single water molecule defines a body. We found that 3-AMOEBA yielded small errors of 0.5-2.0% in regards the total polarization energy, ostensibly indicating acceptable convergence of the MBE. However, convergence of the MBE w.r.t. energies does not correspond to convergence with respect to forces nor the internal virial.[69] Further investigation revealed that the accuracy of forces and virial contributions dramatically improved when the induced dipoles are "embedded" such that the induced dipole sites of the subsystem respond within a much larger permanent electrostatic environment.[69] The electrostatic embedding framework inspired a tractable method of ameliorating the slow convergence of the force due to polarization under the MBE, namely, by defining the "body" in the MBE as a cluster of water molecules rather than a single water molecule and, in turn, greatly reducing the error in the forces when truncating the MBE at the 3-body level.[69] Given the promise of the iAMOEBA and 3-AMOEBA models, in Section 4.2 we show that an order of magnitude improvement in computational speed over the current OpenMP implementation in TINKER can be gained by going to a hybrid OpenMP/MPI parallel strategy.

Section 2.2 QM/MM with Polarizable MM using ONETEP and TINKER

Polarizable force fields such as AMOEBA are expected to improve the description of chemically diverse environments compared to traditional fixed point-charge models. Nevertheless, studies on the use of AMOEBA as the embedding for QM fragments in QM/MM calculations are still very limited due to the complexity of handling the mutual polarization between the two regions. We set out to develop a

consistent model for the coupling of DFT and the AMOEBA polarizable force field using the ONETEP linear-scaling DFT formulation[70] and the TINKER implementation[64] of AMOEBA. Electronic degrees of freedom in ONETEP are described using strictly localized, non-orthogonal generalized Wannier functions, expanded in an underlying p-sinc basis[87]. The MM subsystem employs unmodified AMOEBA electrostatics, i.e. atom-centered, permanent, point multipoles up to quadrupoles and atom-centered, induced point dipoles, with suitably damped and scaled Coulombic interactions[60, 88]. Here, damping refers to a distance-based smoothing of interaction potentials involving induced dipoles, while scaling serves to eliminate or attenuate electrostatic interactions between sites interacting through bonded (valence) terms. The usual mutual polarization scheme is used, with induced dipoles determined self-consistently through an iterative procedure.

In the proposed QM/MM model the polarization between the QM and MM subsystems is fully mutual[89]. This is accomplished by allowing the electrostatic potential of the MM environment (permanent multipoles and induced dipoles) to polarize the electronic charge density through its inclusion in the QM Hamiltonian and in gradients of the total energy with respect to the density matrix. The MM subsystem is in turn polarized in response to an auxiliary representation of the QM charge density (electronic and ionic) in terms of atom-centered, point multipoles up to a quadrupole. This auxiliary representation, hereafter termed QM*, is obtained via an extension to the Gaussian Distributed Multipole Analysis (GDMA)[90] approach to non-orthogonal localized orbitals. The extension that we have developed is a density-fitting scheme[91] that uses truncated spherical waves as an auxiliary basis.

Polarization of the MM subsystem is assumed to occur instantaneously in response to the permanent multipoles that comprise QM*. That is, at every iteration of the total energy minimization, the induced dipoles in the MM subsystem are determined self-consistently by minimizing a separate classical Hamiltonian wherein the QM* multipoles are clamped. The induced MM dipoles are obtained within TINKER, with the QM* multipoles serving as an inactive embedding region. The converged MM induced dipoles, along with permanent MM multipoles, are then included in the QM Hamiltonian in order to recover the polarization response of the QM subsystem, which takes place within ONETEP. The fulfillment of the SCF condition for the induced MM dipoles removes the need for dipole response terms in the derivatives of the QM/MM polarization energy with respect to the density matrix.

In point-dipole polarizability models, a suitable damping of interactions involving induced dipoles is crucial for avoiding the so-called polarization catastrophe. A successful QM/polarizable-MM scheme must also carefully address damping to avoid similar artifacts, e.g. the ill-conditioned interaction of an MM point multipole with a surrounding QM charge density represented on a Cartesian grid with

necessarily finite resolution. In our model the interaction between QM and *induced* MM dipoles leverages the QM* representation, i.e. it is described as a Thole-like damped interaction of point multipoles, which is fully consistent with the AMOEBA model. The QM* contributions obtained from DMA are atom-partitioned, allowing the use of classical polarizabilities for the QM subsystem in the Thole damping expressions. Gradients of this interaction energy with respect to the density matrix elements have been derived and implemented. Although the QM* representation is found to be an excellent approximation of the original QM charge density, its accuracy is nevertheless finite, requiring that the presence of the intermediate step QM→QM* needs to be accounted for in all gradient expressions.

The interaction between QM and *permanent* MM multipoles is not damped, analogous to the AMOEBA model, which does not damp permanent electrostatics. In the absence of damping it is advantageous to forgo the QM* representation, instead obtaining the corresponding energy by integrating the potential of MM multipoles with the full QM density, and thus avoiding charge penetration errors that afflict the point multipole model. Singularities are eliminated by using extremely short-range Thole-like smearing of the multipole potential for grid points in the immediate vicinity (~1a_0) of point multipoles. Validation tests have been undertaken to ensure that the resultant energies are practically insensitive to the details of this smearing.

All strictly-MM energy terms are calculated within TINKER (bonded interactions, MM/MM van der Waals interactions, MM/MM permanent and induced electrostatics), and all strictly-QM energy terms are calculated within ONETEP (DFT energy, empirical dispersion correction)[89]. We find the inclusion of cross-system (QM/MM) van der Waals terms is necessary to obtain a physically sound description of QM/MM interactions, with the repulsive term that implicitly accounts for exchange repulsion serving a crucial function of balancing strong electrostatic attraction between molecules close to the QM/MM boundary. In the current model QM/MM van der Waals interactions are modeled in TINKER, in the same manner as their MM counterparts – except for a suitable adjustment of the slope of the repulsive wall, but with no change to the position of the minimum or its depth.

Our findings indicate that the proposed approach is able to model accurately the embedding of a DFT subsystem within a classical, polarizable environment[89]. This is of practical significance, because it allows chemical accuracy to be obtained with QM regions smaller than those used in fixed point-charge electrostatic embedding schemes, and is vastly superior to neglecting embedding entirely (i.e. performing DFT on a truncated system), with the latter suffering from additional issues beyond its high computational cost[92]. To illustrate this point, Figure 1 presents a test case of a single diphenylhydramine solute in 330 water molecules for which we calculate a reference DFT benchmark of the interaction energy between the

solute and all 330 solvent molecules. We then test whether the benchmark interaction energy can be reproduced with a smaller QM region under the following embedding schemes: all QM with no embedding (truncation), QM with fixed partial-charge electrostatic embedding, QM with GAFF embedding (TIP3P fixed partial charges and Lennard-Jones QM/MM interactions), and QM with mutually polarizable AMOEBA embedding; in all three embedding schemes we keep the total number of water molecules constant at 330. Additionally, we show fully classical results obtained with GAFF and AMOEBA, where we progressively *add* water molecules around the solute.

With no embedding as many as 310 QM waters around the solute are needed to converge the energy to within 1 kcal/mol of the benchmark. Using non-polarizable embedding the accuracy is achieved by retaining 140 water molecules described at the DFT level, improving to 35 once GAFF QM/MM van der Waals interactions are included; whereas with mutual polarization embedding, the number of DFT water molecules is further reduced to 25 DFT water molecules. The fully classical AMOEBA description (green circles) is able to reproduce the interaction energy accurately and tracks the DFT result closely, while the GAFF description suffers from short- and long-range error in excess of 10 kcal/mol. Further validation will be performed to assess the predictive power of this model for calculating reaction energies in solvent. Future theoretical work will focus on three main areas: 1) derivation of polarization energy gradients for the more general case of *in situ* optimized local orbitals, 2) improving the description of van der Waals interactions and implementing bonded terms between the QM and MM subsystems, and 3) investigating more refined models for QM/MM electrostatic interactions.

Section 2.3 QM/MM with Polarizable MM using LibEFP and Q-Chem

LibEFP is an open source library that implements the Effective Fragment Potential (EFP) method, a first-principles-based force-field[93]. While LibEFP was originally designed for EFP calculations[22, 94-95], it can be naturally extended to support the AMOEBA force field in the Q-Chem program[71] for QM/MM calculations. Again, we are interested in a mutual polarization model wherein the contribution of AMOEBA's permanent multipoles and induced dipoles are included in the QM Hamiltonian, while the electric field of the QM back-polarizes the AMOEBA region. The implementation of the AMOEBA force field in LibEFP involved (a) preparing and parsing fragment potential data files ("efp" files) using AMOEBA force field parameters such as permanent multipole moments, atomic polarizabilities, polarization damping factors, vdW parameters and parameters for bonded interactions; (b) using the local frame to obtain the AMOEBA permanent multipole moments from values for the reference frame; (c) adding the van der Waals 7-14 potential used in the AMOEBA force field; (d) including the contributions

from the bonded terms, which are not included in EFP calculations, where each fragment has a frozen geometry; and (e) evaluating the permanent and induced electrostatics of AMOEBA.

The AMOEBA permanent and induced electrostatics can be handled within LibEFP for both non-periodic and periodic systems. For non-periodic systems, the LibEFP code was extended to accommodate the damping scheme and scalar atomic polarizabilities within the AMOEBA force field. For periodic systems, the particle mesh Ewald (PME) method[6, 96], which partitions the computation between the direct space (for short-range components) and the reciprocal space (for long-range components), was added to the LibEFP library. In the direct-space, efficient evaluation of the electrostatic interactions between permanent multipole moments was achieved via a formulation in terms of spherical harmonics[97].

With this implementation in place, the AMOEBA force field in its original formulation can be readily used in QM/MM calculations together with density functional theory (DFT), perturbation theory, and other wavefunction-based theories within the Q-Chem software package. Our current implementation of DFT/AMOEBA calculations is based on a "double" self-consistent field optimization scheme, i.e. the total energy of the QM/MM system is minimized with respect to both electron density from QM atoms (outer loop) and induced dipoles on MM atoms (inner loop), which is different from the gradient-based approach adopted by the ONETEP/TINKER implementation introduced above. This new capability will enable us to predict QM/AMOEBA-corrected hydration free energies, solvatochromism, and many other physical properties. While the AMOEBA force field as implemented within LibEFP has so far been tested only in conjunction with the Q-Chem software package, little effort will be required to get it working with Psi4[98], NWCHEM[99] and other *ab initio* electron structure packages.

Our preliminary results using a single AMOEBA water as an MM fragment (considering permanent and induced electrostatics only, i.e. Coulomb electrostatic embedding) indicate the necessity of smearing the monopoles that correspond to the valence electrons on MM atoms to describe the effect of forward (MM polarizes QM) polarization correctly. One illustrative example is given by Figure 2, where the response of the electron density on the QM water (proton donor in the water dimer) under the polarizing of the AMOEBA water (proton acceptor) is largely improved when the point monopoles that represent valence electrons on the latter are replaced by spherical Gaussian functions[100] with widths that are fine-tuned to reproduce the polarization effect owing to the true QM charge distribution (converged electron density with nuclear charges) of the "environment" water molecule. It should be noted that the smearing of monopoles is only applied when evaluating the contribution of AMOEBA fragments to the QM Hamiltonian, i.e., it does not affect the interactions between MM molecules. Future work may involve developing the parameters or even improving the model for vdW interactions between QM and

MM regions, especially for the repulsive part of the potential (corresponding to Pauli repulsion between the interacting fragments), since it is likely to modulate the QM electron density significantly.

SECTION 3. NEW METHODS FOR ADVANCED POTENTIAL ENERGY SURFACES

3.1. New Approaches to Solving Mutual Polarization

The mutual polarization in Eq. (1) is usually solved for iteratively at each time step in a simulation using methods such as successive over-relaxation (SOR)[63], preconditioned conjugate gradient (PCG)[101], or direct inversion in the iterative subspace (DIIS)[102] methods. The typical trade off among these methods is that they are either expensive (such as SOR) or they speed up convergence through retention of a history of converged mutual dipoles. In molecular dynamics calculations, this history retention leads to poor energy conservation through a systematic energy drift. An alternative approach is to dynamically evolve the polarization degrees of freedom alongside the 'real' atomic system in the form of an extended Lagrangian (EL).[103-105] However, EL formulations typically suffer from numerical stability issues that can only be addressed by decreasing the size of the time step, which, for obvious reasons, is undesirable.. In all cases it should be emphasized that solutions to the mutual polarization are approximate and will lead to numerical accuracy problems that must be controlled.

A number of new approaches were introduced in 2015 that offer a better balance between accuracy, speed, and energy conservation, which we can by classified as new approximate models vs. new methods to solve for the mutual polarization response. In the former category the iAMOEBA model solves the direct polarization exactly[68], while the 3-AMOEBA model using single water molecule bodies approximates the N-body mutual polarization by summing over small dimer or trimer fragments whose polarization can be solved analytically through matrix inversion[20]. Another example is the Extrapolated Perturbation Theory Optimization (ExPT-Opt3)[106], which is also formulated as an analytical but approximate model to AMOEBA's mutual polarization. ExPT begins with the direct polarization model and then sets up a perturbation theory approach to obtain a mutual polarization energy functional that approximates the full induced dipole variational coupling. In practice the perturbation theory is replaced with an extrapolation scheme that attempts to fit the infinite series solution, , and the quality of said fit depends on the system being studied. Because these are approximate models that discard part of the full AMOEBA mutual polarization solution, all three approaches require some reparameterization of the AMOEBA polarization parameters to recapture any missing mutual polarization or to address the system-dependence of the model.

By contrast, there are now significantly better methods for solving mutual polarization of the standard AMOEBA model directly, beyond the conservative SOR procedure that was the default solver in the TINKER software platform for two decades. This changed in 2013 when the SOR was replaced by the more efficient SCF iterative methods like preconditioned conjugate gradient (PCG) in TINKER[64] and direct inversion in the iterative subspace (DIIS) for classical polarization calculations in TINKER-HP[107-108]. These offer significant computational speed-ups by greatly reducing the number of SCF cycles, and can be further supplemented by a predictor that extrapolates an initial guess for the SCF solver from previous SCF solutions at previous time steps. However while a predictor may seem like a straightforward method to reduce SCF cycles, it destroys the underlying time reversibility thereby affecting both energy conservation and convergence. In general a non-predictor method will be able to maintain energy conservation at a lower convergence threshold than a predictor method due to its time-reversibility. For example, AMOEBA water in the NVE ensemble with a PCG solver requires approximate convergence thresholds of 10^{-6} RMS D with a predictor and 10^{-3} RMS D without one.

This tradeoff between energy conservation, time-reversibility, and convergence threshold can undermine some of the usefulness of a traditional predictor. Here we highlight the inertia-restrained extended Lagrangian/self-consistent field (iEL/SCF) method that overcomes these problems[72]. The iEL/SCF method eliminates the predictor used in PCG and DIIS and allows for the use of a much relaxed convergence criteria for SCF iteration. In general iEL/SCF employs time-reversible propagation in extended Lagrangian methods in conjunction with numerically stable SCF iteration. In the spirit of Niklasson *et al*. who originally developed the approach for iteratively finding the ground state density in *ab initio* molecular dynamics simulations [109-113], we define an extended Lagrangian given by Eq. (1) for the induced dipoles used to describe classical polarization:

$$\mathscr{L}_{hybrid}^{dipole} = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\vec{r}}_i^{\,2} + \frac{1}{2}\sum_{i=1}^{N} m_{\mu,i}\, \dot{\vec{\mu}}_i^{\,2} - U\left(\vec{r}^{\,\boldsymbol{N}}, \vec{\mu}_{SCF}^{\,\boldsymbol{N}}\right) - \frac{1}{2}\omega^2 \sum_{i=1}^{N} m_{\mu,i}\left(\vec{\mu}_{SCF,i} - \vec{\mu}_i\right)^2 \qquad (6)$$

where $\vec{\mu}_{SCF}^{\,\boldsymbol{N}}$ represents the set of all converged physical induced dipoles. In Eq. (6) we have also introduced another set of induced dipoles, $\vec{\mu}^{\boldsymbol{N}}$, which are the initial guesses to the iterative solution of $\vec{\mu}_{SCF}^{\,\boldsymbol{N}}$. This auxiliary set of induced dipoles is restrained to stay near the true self-consistent values via the final term in Eq. (6) which is a harmonic function characterized by a fictitious mass, $m_{\mu,i}$, and a frequency,ω; the latter is a universal parameter that determines the curvature of the harmonic well. Applying the Euler-Lagrange equation of motion to Eq. (6) in the limit that $m_{\mu,i} \longrightarrow 0$ yields the equations of motion for atomic centers and induced dipoles,

$$m_i \ddot{\vec{r}}_i = -\left.\frac{\partial U\left(\vec{r}^{\boldsymbol{N}}, \vec{\mu}^{\boldsymbol{N}}_{SCF}\right)}{\partial \vec{r}_i}\right|_{\vec{\mu}^N} \tag{7a}$$

$$\ddot{\vec{\mu}}_i = \omega^2\left(\vec{\mu}_{SCF,i} - \vec{\mu}_i\right) \tag{7b}$$

Eq. (7a) shows that equations of motion for the atomic centers are propagated in the usual way, except that the PCG iterative solution to determine $\vec{\mu}^N_{SCF}$ now uses an initial guess that is propagated by the auxiliary electronic degrees of freedom in Eq. (7b). We integrate both equations of motion using time-reversible velocity Verlet integration [114], and we chose ω to be $\sqrt{2}/\Delta t$, where Δt is the time step, which we set to 1 fs.

We have previously shown [72] that this straight adaptation of Niklasson's method is insufficient for both for decreasing computational expense and maintaining numerical stability due to problems that arise from resonances that make the auxiliary dipoles increasingly poor initial guesses for the SCF solutions of the physical induced dipoles. More specifically, we showed that the auxiliary dipoles evolve on a much faster timescale than their physical counterparts owing to the optimal choice of their characteristic frequency, $\sqrt{2}/\Delta t$, and their direct coupling in the auxiliary potential (Eq. (7b)) leads to corruption of their dynamics. To address this problem, we introduced "temperature" control on the auxiliary set of dipoles, using both Berendsen[115] rescaling and time-reversible Nosé-Hoover chains[9-10] as thermostating methods, much as is done in Car-Parrinello *ab initio* MD calculations[116-117]. The set point pseudo-temperature of the auxiliary system can be determined with an equipartition argument applied to the harmonic restraining forces experienced by the auxiliary dipoles[72].

The iEL/SCF method shows much better energy conservation over a wide range of convergence levels for SCF iteration, especially compared to typical SCF methods that use a predictor as the CG-SCF method implemented in TINKER. Furthermore, since iEL/SCF can employ a looser SCF convergence criterion, fewer SCF iterations are required, typically half as many. Moreover, all physical properties predicted using standard SCF with a tight convergence criterion are well reproduced by iEL/SCF with a loose convergence criterion. These include both equilibrium thermodynamic quantities, such as the average molecular dipole moment and average potential energy, as well as dynamical properties such as the diffusion constant (Table 1).

Section 3.2 Increasing the Molecular Simulation Timestep for AMOEBA

Most properties of interest in a condensed-phase system are connected to low-frequency, long time-scale phenomena. However, converging such properties generally requires many small time steps the size of

which is dictated by the simultaneous presence of the highest frequency motions in the system. The fundamental problem posed by the need for a small time step corresponding to the fastest evolving degrees of freedom can be addressed via multiple-time step algorithms such as r-RESPA[66]. This approach is based on the recognition that forces evolving on different time scales may be evaluated at different time intervals. More specifically, the forces due to the faster covalent terms are evaluated more frequently and the more slowly evolving forces due to non-covalent interactions evaluated less frequently, all within a symplectic framework. The computational savings garnered derives from the fact that the cost of the force calculation in classical MD is dominated by the nonbonded contributions, a cost that is substantially lowered when the number of such force evaluations is thus reduced.

Standard multiple time-step (MTS) approaches are limited, however, by resonance phenomena, wherein motion on the fastest time scales poses a typical limit on the size of the time step that may be used for the calculation of the slowly evolving non-covalent forces. In atomistic models of biomolecular systems, for example, the largest time step is around 3-5 fs. Previously, Tuckerman and co-workers introduced an isokinetic extended phase-space algorithm[118] and its stochastic analog[119] that eliminate resonance phenomena through a set of kinetic energy constraints. Using such constraints, a time step of nearly 100 fs could be employed for the slowly evolving forces in a simulation of liquid water using a fixed-charge flexible model.

Tuckerman and coworkers have been able to extend the stochastic resonance-free approach, termed stochastic isokinetic Nosé-Hoover (RESPA) or SIN(R), of Leimkuhler *et al* [119] to polarizable models formulated in terms of fluctuating induced dipole moments[73]. Note that SIN can be used as a canonical sampling method with or without a RESPA[66] MTS decomposition. Typical MTS algorithms for fixed-charge force fields involve decomposing the forces into bonded forces, short-range non-bonded forces, and long-range non-bonded forces. Although the forces in polarizable models cannot be explicitly decomposed in this fashion, the concept can, nevertheless, be adapted to the polarizable case by introducing a short-ranged induced dipole moment calculation based on a spherical cutoff, r_{cut}, which is then used to construct induced dipole forces in the short-range steps. These short-range dipoles are given by the equation

$$\mu_{i,\gamma}^{(d,S)} = \alpha_i \left(\sum_{j}^{13N} T_{ij,\gamma} M_j^{(S,d)} + \sum_{k, r_{ik} \le r_{cut}}^{3N} T'_{ik,\gamma\delta} \mu_{k,\delta}^{S} \right) \qquad \gamma, \delta = x, y, z \tag{8}$$

This is then corrected in long-range force steps by performing a full calculation of the induced dipole moments and subtracting the short-range contribution, where the full dipole calculation follows the usual prescription of Eq. (3). All pairwise forces are decomposed into short- and long-range contributions in the

standard fashion using a spherical cutoff, and reciprocal-space forces coming from Ewald summation are evaluated with the long-range forces, forming a scheme we call RESPA1-pol. This approach has been implemented in the TINKER7 software package, and in MD simulations of 512 AMOEBA water molecules in a periodic cubic box of length 24.85 Å, it is found that the time step for the expensive long-range calculations can be increased to 120 fs with no degradation in the equilibrium properties. This is illustrated for the oxygen-oxygen and oxygen-hydrogen RDFs in Figures 3a and 3b. Importantly, we obtain computational speedup factors ranging between 10 and 25, depending on the choice of simulation parameters (see Figure 3c). Moreover, although not entirely expected, the SIN(R) algorithm accurately reproduces diffusion constants, as shown in Table 2.

Section 3.3 New Formulations of Electrostatic Multipolar Interactions

Developing software that implements multipolar electrostatics interactions for molecular dynamics engines in a parallel, efficient and sustainable manner presents many challenges. Despite the wealth of theoretical and methodological work only a few software projects have dealt with these challenges including DL_MULTI[120], TINKER[64], Amber[2], and CHARMM[3]. It is worth noting that DL_MULTI implements the multipolar expansion up to hexadecapole moments using a formulation in terms of spherical harmonics whereas TINKER limits the expansion to quadrupole moments on point centered multipoles (within flexible molecules, i.e. no rigidification restrictions) using a Cartesian formulation. Both software packages rely on replicated data parallelization, which is not memory distributed and is known to suffer from memory overheads with large model system sizes and prohibitive communication overheads on anything beyond moderate processor counts (~100).

The GDMA program[90] used to define multipole electrostatics allows for generation of multipoles up to tenth order if it is desired to include multipoles beyond quadrupoles as is currently the practice for AMOEBA development. One of the main new contributions to the DL_POLY[121] and CHARMM software packages is the generalization of multipolar particle mesh Ewald (PME) to arbitrary order that builds on previous work on fixed charges, dipoles, quadrupoles, and hexadecapoles. It is hoped that these new methods will lower the barrier for using higher order multipoles in future developments of other advanced potential energy models and their software implementations.

The key idea of PME is to approximate the structure factor on a uniform grid in 3 dimensions that fills the simulation cell. We have derived a closed form formula for arbitrary order multidimensional derivatives of the product of three B-splines which is required for the extension of the reciprocal space PME to arbitrary order, as well as for the stress tensor needed for constant pressure simulations, and

provided a simple procedure for the B-spline computation using particle mesh Ewald[74]. That study also provided recurrence relations for the Coulomb sum, force-shifted Coulomb, Coulomb sum with distance dependent dielectric, and reaction field, which allows for a simple implementation of permanent multipoles up to arbitrary order with potentials other than the Ewald sum. While there has already been work in this direction by Nymand and Linse[122] and Brommer et al.[123], Boateng and Todorov provided an alternative recurrence relation for the real space kernel of the Ewald sum[74]. Furthermore, it was shown in their work that there is a Cartesian recurrence relation that has cubic scaling in the order of multipole used, which is equivalent to the scaling of a recently developed spherical harmonic version of the real space that also achieves cubic scaling by Simmonett et. al. and implemented in CHARMM[97]; both improve on the original scaling of the Boys recurrence in the original PME formulation[7] which is quartic in the multipole order. The development of the new Cartesian multipolar approach has now been integrated efficiently within DL_POLY_4's domain decomposition framework to utilize the performance advantages of SPME[96, 124,125] for the Ewald case.

Section 3.4 Approaching the basis set limit for DFT calculations with advanced functionals

The performance of QM/MM methods is limited not just by errors in treating the QM/MM boundary or limitations of the force field, but also by the inexactness of the QM model itself. Fortunately, the statistical errors of standard density functionals like PBE-D3[126-127] can be reduced by roughly a factor of two for thermochemistry (TC) and non-covalent (NC) interactions by recently developed density functionals, such as ωB97X-V[128] and B97M-V[129]. The latter has the computational advantage of not requiring exact exchange (which necessarily leads to larger errors for self-interaction sensitive problems).

It is important to emphasize that such impressive improvements in accuracy are only achieved when very large atomic orbital basis sets are used (ideally beyond triple zeta, for instance the quadruple zeta def2-QZVPPD basis)[130]. However, in conventional quantum chemistry, the cost of approaching the one-particle basis set limit is very high, scaling at least as $O(n^3)$ with the number of functions per atom, due to Fock matrix element evaluation and diagonalization costs. Therefore practical DFT calculations are typically performed with smaller basis sets, and it is an important challenge in quantum chemistry to reduce the cost of large basis calculations. Indeed, DFT based ab initio molecular dynamics calculations of pure water and aqueous solutions using fully converged discrete variable representation basis sets show that both structural and dynamical properties are sensitive to the size of the basis set[131-137]. Examples of existing approaches that help defray the computational cost are the use of resolution of the identity (RI)

approximations for the Coulomb and exchange interactions[138-141], which are nowadays standard, as well as the dual-basis SCF methods.[142-143]

In order to approach the basis set limit with potentially lower computational effort, we have been working on a new SCF scheme using a minimal adaptive basis (MAB) consisting of functions that are variationally optimized on each atomic site.[75] We shall review this approach briefly, which aims to reproduce the accuracy of large-basis SCF calculations. It involves a sequence of four steps to produce the final energy.

The first step is to project the small 6-31+G(d) basis into the large target basis, and performing an SCF calculation in this projected basis (of the small basis dimension) in order to obtain a reference one-particle density matrix, $\mathbf{P}_{ref}$. Relative to full SCF, this offers a cost reduction of $(N/P)^3$ where N is the target basis rank and P is the rank of the 6-31+G(d) basis. The second step is then to find the MAB by minimizing a judiciously chosen energy-like function:

$$L = \mathrm{Tr}\left[\mathbf{RSP}_{ref}\mathbf{SRF}\right] \tag{9}$$

This function depends only on the converged Fock matrix, $\mathbf{F}$, corresponding to $\mathbf{P}_{ref}$, so iterative minimization involves only linear algebraic operations. Eq. (9) is minimized with respect to the elements of an atom-blocked transformation, $\mathbf{B}$, from the large target basis to the MAB representation, and this atom-blocking reduces the scaling of many linear algebra steps from cubic to quadratic. $\mathbf{B}$ determines the span of the minimum basis, $\mathbf{R}$, through $\mathbf{R} = \mathbf{B}\sigma^{-1}\mathbf{B}^{\mathrm{T}}$, where $\sigma = \mathbf{B}^{\mathrm{T}}\mathbf{SB}$ and $\mathbf{S}$ is the atomic orbital overlap matrix.

The third step involves performing an SCF calculation in the MAB, which yields a cost reduction of $(N/M)^3$ relative to the target basis SCF, where M is the rank of a minimal basis. Finally, in the fourth step, the desired accuracy is obtained by applying a perturbative correction (PC) to the SCF/MAB solution. For a pure functional like B97M-V, the PC simply uses the converged MAB Fock matrix, similar to dual-basis SCF methods[142-143]:

$$E_{tot} = E_{MAB} - \sum_{ia} F_{ai}{}^2 / \left(\varepsilon_a - \varepsilon_i\right) \tag{10}$$

The sum runs over occupied orbitals, i, and virtual orbitals, a, and ε are semi-canonicalized Kohn-Sham eigenvalues. This expression requires solution of a single large system of linear equations, giving a speed-up relative to conventional SCF that is greater than the typical iteration count (between 10 and 20).

Compared to exact SCF results using the modern B97M-V density functional, using this MAB-SCF (PC) approach with the same target basis set produces < 0.1 kcal/mol RMS errors for tested TC data, and less than 0.05 kcal/mol for the NC data. Hence the performance of modern density functionals near

the basis set limit can be quite faithfully reproduced. As an illustration of the resulting basis set convergence, Figure 4 shows the results of accuracy tests on 5 NC datasets using B97M-V, using the MAB approach and conventional SCF in the def2-QZVPPD basis; the results are also compared to the best available reference data.

Two points emerge clearly from these tests. First, the B97M-V density functional yields quite small RMS deviations against the reference data. To see how small, consider that the data points in H2OBind8, for instance, represent the binding energies of water hexamers that are each 40-50 kcal/mol in magnitude. In terms of accuracy, these results are close to the best that is currently possible with DFT. Second, the difference between the conventional large-basis calculations and the MAB-based calculations is a factor of roughly 10 times smaller than the intrinsic error associated with the B97M-V density functional. Thus the MAB-based model, indeed, successfully approaches the basis set limit, as was our goal.

Finally, we should note a few caveats and limitations. First, molecules that contain hypervalent atoms require extra MAB functions beyond a standard minimum basis dimension. This is accommodated by a threshold-driven adaptive scheme. Second, the excellent accuracy shown for pure functionals can only be matched for hybrid functionals if a more sophisticated (and computationally twice as costly) PC is used at twice the computational cost. Third, and most important, our present implementation in Q-Chem[71] is a pilot implementation which does not yet optimize the savings that are possible for matrix element evaluation, although it is nearly optimal with regard to the linear algebra effort. In conclusion, with further improvements in its implementation, MAB-SCF (PC) could be a promising low-cost substitute for its conventional counterpart to viably approach the basis set limit of modern density functionals.

SECTION 4. SOFTWARE IMPROVEMENTS

4.1 OpenMP Performance Optimization for AMOEBA in TINKER

A primary objective of the consortium work has been to improve performance on CPUs with regard to the canonical implementation code, TINKER, starting with improvements to the existing OpenMP framework which is most suitable for running on in-house commodity clusters. In this case, original versions of the reference implementation of AMOEBA in the TINKER software were based on explicit Cartesian multipole formula for the energy and gradient charge-charge multipole interactions up to quadrupole-quadrupole multipole interactions. Thole damping of induced dipole interactions was handled as a special case, and computed in-line as part of the energy and gradient (force) calculations. In an attempt to streamline the TINKER code and make it easier to add modifications such as explicit charge penetration

effects[52-53], the latest version of TINKER computes and stores the electrostatic potential, electric field, electric field gradient and electric field Hessian due to permanent multipole moments, and corresponding quantities for the induced dipoles, during the initial solution of the polarization equations, Eq. (3). This greatly simplifies the subsequent calculation of the energy and gradient. For example, the polarization energy is then just the dot product of the induced dipoles with the electric field. In addition this refactoring allows easy testing of alternative short-range damping functions since changes are only needed in the field and its gradients, without any alteration of the energy and gradient code.

4.2 OpenMP/MPI Performance Optimization in TINKER and TINKER-HP

Recently we have implemented a parallel hybrid MPI/OpenMP replicated data (atom-decomposition) scheme that is most suitable for running iAMOEBA in TINKER on HPC platforms having a fast node interconnect. In this case we sought to reduce the computational expense by pre-calculating and saving the tensor elements corresponding to the permanent field and field gradient. For instance, for a multipole site i interacting with an induced dipole site j, we would have the following derivative terms of the permanent field that would need to be saved:

$$dE_{ij,xx} \equiv \frac{\partial}{\partial x_i}\left(\sum_k^{13} T_{ij,k,x} M_{i,k}\right) \qquad dE_{ij,xy} \equiv \frac{\partial}{\partial y_i}\left(\sum_k^{13} T_{ij,k,x} M_{i,k}\right) \qquad dE_{ij,xz} \equiv \frac{\partial}{\partial z_i}\left(\sum_k^{13} T_{ij,k,x} M_{i,k}\right)$$

$$dE_{ij,yx} \equiv \frac{\partial}{\partial x_i}\left(\sum_k^{13} T_{ij,k,y} M_{i,k}\right) \qquad dE_{ij,yy} \equiv \frac{\partial}{\partial y_i}\left(\sum_k^{13} T_{ij,k,y} M_{i,k}\right) \qquad dE_{ij,yz} \equiv \frac{\partial}{\partial z_i}\left(\sum_k^{13} T_{ij,k,y} M_{i,k}\right)$$

$$dE_{ij,zx} \equiv \frac{\partial}{\partial x_i}\left(\sum_k^{13} T_{ij,k,z} M_{i,k}\right) \qquad dE_{ij,zy} \equiv \frac{\partial}{\partial y_i}\left(\sum_k^{13} T_{ij,k,z} M_{i,k}\right) \qquad dE_{ij,zz} \equiv \frac{\partial}{\partial z_i}\left(\sum_k^{13} T_{ij,k,z} M_{i,k}\right) \tag{11}$$

all of which would need to be saved. The direct contribution to the polarization energy is calculated simply as the dot product of induced dipole and the saved permanent field. The direct polarization contributions to the gradient and internal virial are then calculated by iterating over the saved tensor elements. The resulting tensor method, on its own, was found to be a computationally advantageous method for the direct polarization calculation using the iAMOEBA model, in particular, which is finding increased popularity within the water community.

Hybrid MPI/OpenMP parallel timings in nanoseconds per day were extrapolated from short molecular dynamics simulations of 100-2000 time steps using the velocity Verlet integrator, a 1 fs time step, a 9.0 Å vdW cutoff, and a 7.0 Å real space cutoff for particle-mesh Ewald (PME) electrostatics, and using both small numbers (384 or 768) up to large numbers (3072 to 6144) of cores on system sizes ranging from 1600 to 288,000 water molecules. The reference TINKER7 code is parallelized in the

shared-memory regime with OpenMP, so all timing comparisons with the reference TINKER7 were obtained using 12 OpenMP threads that gave the best speedup. Table 3 presents the timings and speedups for the hybrid MPI/OpenMP implementation, with the tensor method achieving a factor of ~4.5 for low numbers of cores and up to a factor of ~7.9 with larger numbers of cores.

The very nature of the MBE approximation admits a trivial parallel implementation for the approximate mutual polarization model 3-AMOEBA, as the polarization energy, gradient, and virial of the subsystems are independent of one another. Again, we use a hybrid MPI/OpenMP approach and a load-balancing scheme is implemented that ensures that the work of calculating the polarization energy, gradient, and virial is distributed as evenly as possible among the MPI tasks. Table 4 shows that the approximate 3-AMOEBA model is faster by factors of 2 to 11 compared to the parent AMOEBA potential, depending on the fragment size that defines a body n, and the total system size N. The 3-body approximation by its very nature imposes more work but trivial parallelization compared to the AMOEBA model. In the process of improving the accuracy of polarization gradients under the 3-body approximation, we have fortuitously ameliorated this problem by concomitantly increasing the size of the fragment, thereby reducing the number of fragments. Nonetheless, the computational speedup of our method relies on the simultaneous calculation of 1-, 2-, and 3-body contributions to the energy, gradient, and virial, with each MPI task dedicated to a single subsystem calculation, thereby necessitating up to 3600 cores. In an era of exascale computing, the 3-AMOEBA would be highly suitable as a refactored form of the AMOEBA potential. In contrast to 3-AMOEBA, the direct polarization model of iAMOEBA does not suffer from a curse of numbers of independent calculations and computational speedups over the reference TINKER7 implementation may be realized with a few hundred cores.

The recently developed TINKER-HP software provides a new implementation of AMOEBA for high-throughput production calculations on CPUs. TINKER-HP is a distributed memory version of the core AMOEBA energy and force functions using MPI parallelization. Previous results have shown good scaling on the polarizable multipole PME electrostatics calculations that are a major computational bottleneck in AMOEBA simulations.[52] The TINKER-HP code has now been extended to perform full AMOEBA molecular dynamics and minimization calculations. Excellent scaling performance is observed for AMOEBA MD simulation across thousands of cores for large systems such as the 3.5 million atom solvated ribosome complex. Once testing is complete, the code for TINKER-HP will be released to a public GitHub site, and we anticipate it may become the implementation of choice for AMOEBA calculations on large-scale, high-performance distributed computer systems.

4.3 OpenMM for GPU-Based Calculations with AMOEBA

The OpenMM software[5] (http://openmm.org) developed by the Simbios Center at Stanford University is highly optimized to perform MD simulations on graphical processing units (GPUs). It is part of the Omnia software suite (http://omnia.md), a collection of interoperating tools for biomolecular simulation. It can be used either as a standalone simulation package, or as a library to perform calculations within other applications. It has supported the AMOEBA force field since version 3.0, released in 2011.

AMOEBA simulations can be run with OpenMM in several different ways. One option is to use it as a standalone package, writing a Python control script to direct the details of the simulation. Alternatively, a set of scripts is available to automate the preparation and running of simulations (https://github.com/apes-soft/OMM_Amoeba_Scripts). Finally, there is a GitHub repository containing the TINKER interface to OpenMM (with modifications to base OpenMM specific to AMOEBA calculations) at (https://github.com/pren/TINKER-openmm). This allows users to run simulations with TINKER while having the calculations be done by OpenMM on a GPU.

The advantage of running the simulation directly in Python (rather than through TINKER) is that this provides access to a wider range of popular output file formats such as DCD or NetCDF binary trajectory files, as well as established reversible reference system propagator algorithm (RESPA) multi-time stepping integrators[66, 118] that are currently available only through the OpenMM Python interface. Furthermore, it eliminates the requirement that TINKER be installed on the machine used to run the simulation (although at the expense of requiring Python to be available).

The first approach requires using the OpenMM modeling toolkit to parameterize a structure defined in a PDB or PDBx/mmCIF file containing standard biomolecular residues, water, and basic ions. This process is automated through one of the scripts in the aforementioned GitHub repository. This approach is the simplest to execute, as it requires only installing OpenMM (which is done using the *conda* package manager in a simple manner) and executing a preparatory script on a lightly-curated PDB file. It is limited, however, in that it works only for standard amino and nucleic acid residues as well as water and a small number of common monatomic ions.

Applying the AMOEBA force field to systems containing nonstandard residues or nonstandard post-translational modifications requires TINKER to generate a serialized representation that can later be used to run a simulation using an OpenMM-powered Python script. This approach is the most flexible, since it utilizes the de facto reference implementation of AMOEBA to reliably parameterize the target system. However, it also requires users to install not only OpenMM, but also TINKER *and* the TINKER-

OpenMM interface. Since there are no precompiled binaries for this package and development on this functionality is occurring rapidly, it can be difficult to even install the prerequisite software to execute this workflow. Furthermore, OpenMM seems to be more tolerant of variations in PDB files, meaning that this second approach involving TINKER also requires curating the PDB files much more carefully. We have generally found that for systems that are capable of being parameterized entirely using OpenMM, setting up an AMOEBA simulation using the first approach is substantially more efficient.

Table 5 shows the total throughput of AMOEBA MD simulations on a single commodity NVIDIA GTX 980 GPU card via the TINKER-OpenMM interface running in "mixed" precision. Mixed precision means that the calculated forces, the bottleneck in all MD calculation, are evaluated in single precision to provide the maximum speedups, whereas the accumulation of these interactions as well as the actual integration step (update of positions, velocities and accelerations) are taken in either double or fixed precision to improve the accuracy. Table 5 also reports double precision timings in OpenMM generated on the Tesla K80 GPUs. We note the speedup on GPUs running double precision is ~5 times faster, whereas using mixed precision is roughly 10-20 times faster, than that available with double precision versions of TINKER running OpenMP parallelization on traditional multi-core CPU compute nodes. In addition, speed up doubles when using the iELSCF[72] method or ExPT/Opt3[106] model for solving for mutual polarization, and effective timescales could be much longer if the new MTS scheme[73] of Sec. 3.2 was used for this test. The ability to run 100ns simulations in a reasonable time on small to medium-sized proteins opens up the possibility of using AMOEBA to perform a wide range of ligand and drug-binding free energy calculations or long biomolecular simulations in which sampling is important, for example for intrinsically disordered proteins.

SECTION 5. VALIDATION AND APPLICATIONS

Section 5.1. Validation Studies of the AMOEBA Potential

A major objective of the AMOEBA force field since its inception has been to produce a tractable model for rapidly computing physics-based binding energetics in host-guest and protein-ligand systems. Figure 5 shows results from application of the AMOEBA model to the host-guest binding challenge from the SAMPL4 exercise.[144] All of the ligands are organic ammonium cations and they are bound to cucurbit[7]uril (CB7), which is well-known to interact with cations via its rings of carbonyl oxygen atoms. The interior of the CB7 barrel is much less polar than its carbonyl-based portals, and is able to bind hydrophobic alkyl and aryl moieties. The predicted free energies in Figure 5 represent absolute

binding energies computed from molecular dynamics (MD) sampling via a standard double-decoupling method[145] and the Bennett acceptance ratio (BAR) protocol.[146]

Of the roughly 20 prediction sets submitted for the SAMPL4 CB7 host-guest series, the AMOEBA results were at or near the best reported. The AMOEBA results exhibit over binding for multiply-charged guests, *i.e.*, guests 1, 4 and 9 are all dications, while the other guest are all singly-charged. The largest individual error is for guest 5, which is by far the smallest of the 14 and is under bound compared to experiment. Short 1 ns MD windows were used in the BAR free energy calculations, and work to determine the effect of the method and extent of sampling is in progress.[147] AMOEBA has recently been applied to the corresponding host-guest systems from the SAMPL5 exercise, and those results will be made available in due course.

Of course the blind prediction outcomes using AMOEBA in this and other SAMPL exercises[148] provide impetus for improving the AMOEBA force field. Electronic structure calculations have been routinely used to provide training data for the parameterization of molecular mechanics force fields, and can provide useful benchmarks for properties such as interaction energies on molecular systems outside the original training set. Although the overall QM interaction energy is often used to validate a classical force field, the advent of energy decomposition analysis (EDA) affords a new way to independently quantify the individual contributions of several physically meaningful terms out of the QM interaction energy, e.g. permanent electrostatics, polarization, dispersion, and charge transfer[149-151]. Therefore, by comparing the corresponding terms between an EDA scheme and a molecular mechanics potential, one can obtain insight into the strengths and weaknesses of the force field of interest.

Several popular categories of EDA methods include symmetry-adapted perturbation theory (SAPT)[152-154], and variational methods that originate from the Kituara-Morokuma EDA formalism[149, 155-156,157-160], including the absolutely localized molecular orbital (ALMO)-EDA developed by Head-Gordon and co-workers.[161-163] In recently submitted work, we have used the second generation of (ALMO)-EDA[164-166], combined with the high quality ωB97X-V density functional[128], to compare against the energy decomposition of terms in the AMOEBA potential for the water dimer and various simple ion-water dimers[76], an example of which is shown in Figure 6. For those interactions, we showed that the ωB97X-V/def2-QZVPPD level of theory is virtually equivalent to reference CCSD(T) calculations at the complete basis set limit, at a fraction of the computational cost, as CCSD(T) scales as $O(N^7)$. Furthermore, the recent development of the fragment electric field response function model has enabled a meaningful complete basis set limit to be reached for polarization and charge transfer in the second generation ALMO-EDA approach.[164]

Figure 6a shows the total interaction energy for the potassium-water dimer as a function of the K^+-O_w distance, in which it is notable that AMOEBA gives quite good agreement for the equilibrium distance and binding energy when compared to the ωB97X-V benchmark using a def2-QZVPPD basis set.[130] However the breakdown into energy components in Figure 6b shows that while AMOEBA's polarization potential is almost perfect near and beyond the equilibrium position, the primary total electrostatic error resides in the permanent multipole contribution, which is far too unfavorable. This discrepancy in permanent electrostatics could largely be ameliorated through introduction of the charge penetration effect into the AMOEBA model. Nonetheless, Figure 6c emphasizes several aspects of the cancellation of errors at play that can explain the overall excellent result in Figure 6a. First is that while the inclusion of charge transfer does improve the agreement between AMOEBA and EDA's repulsive part of the vdW potential, there is still a remarkable difference in that the wall is still too soft while the dispersion is too favorable compared to ALMO-EDA, which seems to be able to compensate almost perfectly for AMOEBA's less favorable total electrostatics.

We anticipate that high quality EDA approaches may be a particularly fruitful scheme for advanced molecular mechanics potential that, when combined with automated force field parameterization[45, 83-84], may yield next generation force fields with greatly improved accuracy by more correctly accounting for the individual energy components of the total QM interaction. Alternatively EDA results can guide the intelligent inclusion of some effects (e.g. CP), while others are neglected implicitly (e.g. CT) and lumped into a less-softened repulsive wall of the van der Waals potential.

Section 5.2. Large-Scale Applications of the AMOEBA Potential

Similar to the ion-water interaction potentials, force fields for macromolecules are universally constructed from fragments (such as amino acids) whose parameters come from studies of even smaller molecules. In general, one has to check that the "emergent" properties of macromolecules are faithfully represented when the fragments are combined; this may be of special concern for polarizable potentials, where mutual polarization is a non-additive property. Based on advances in previous sections, we have carried out some of the first very long MD simulations using the AMOEBA potential on three small, globular proteins: fragment B3 of protein G ("GB3", 61 residues), hen lysozyme (129 residues) and cyclophilin A (161 residues). Figure 7 shows the time dependence of the root-mean-square (rms) deviation of the backbone atoms from the crystal conformation for the three proteins, comparing AMOEBA with two recent fixed-charge force fields. The primary conclusion is that the folded state of proteins using the AMOEBA

potential is within error bounds of that obtained using current fixed charge force fields, the latter of which have had decades of training and tuning to obtain specific desired results.

We have also applied AMOEBA to the calculation of the electric fields of Cyclophilin A. Cyclophilin A is one of a family of peptidylprolyl isomerases catalysing the interconversion between *cis* and *trans* proline in peptide substrates. Recently, a novel 'electrostatic handle' catalytic mechanism was proposed using molecular dynamics from which snapshots were evaluated with quantum mechanical calculations, whereby the surrounding protein exerts a directionally oriented environmental field on the substrate peptide bond, stabilizing the transition state of the *cis/trans* interconversion (Figure 8).[167,175] Since the AMOEBA force field has previously been shown to accurately recreate the environmental field strengths exerted on an acetophenone probe by a variety of solvents, and has been used to calibrate experimental frequency shifts observed in vibrational Stark effect spectroscopy[14], we investigated the ability of AMOEBA to recreate proposed environmental field strengths in the enzyme cyclophilin A.

Field strengths observed during AMOEBA MD simulations (Table 6) reproduced the trends in field strengths suggested by Camilloni *et al.* from QM calculations extremely well, across both *cis* and *trans* substrate endpoints (Table 6, column 3), across different substrate sequences (Table 6, rows 1-4) and for Cyclophilin B, a different member of the cyclophilin family (Table 6, row 5). In addition, a large field drop was observed in simulations of the R55A mutation in cyclophilin A. This arginine residue, which is highly conserved across members of the cyclophilin family[168], was proposed to be the source of the majority of the environmental field strength (and hence catalytic activity) in the protein. This is consistent with the trends observed in our simulations, and suggests that the complex electrostatic environment of an enzyme active site, and changes therein, can be recreated well by AMOEBA. We continue with further work to validate these observations across multiple enzymes with vastly differing environmental field strengths and catalytic mechanisms.

SECTION 6. CONCLUSIONS

Classical force fields are systematically progressing beyond the well-established but fundamentally limited fixed, atom-centered monopole models. In particular they are starting to adopt more sophisticated descriptions of permanent electrostatics and many-body effects that can allow for a more accurate reproduction of a much broader range of reference data and to make better predictions. However, the greater accuracy introduced by improvements in short range forces that now include fixed multipoles and polarizability, and that are evolving to include charge transfer and charge penetration or explicit quantum mechanical treatments, are revealing challenges in their software implementation on modern hardware

platforms that preclude their widespread adoption by the computational chemistry community. We believe that the results we have summarized on models, algorithms and application studies centered on multipole-based polarizable potentials such as AMOEBA, have reached a landmark on usefulness to the broader computational chemistry community. Now application scientists have an expanded set of capabilities in regards advanced potential energy surfaces in a range of community codes including Amber, Charmm, DL_POLY4, LibEFP, ONETEP, OpenMM, Q-Chem, TINKER, and TINKER-HP. While multipole-based polarizable force fields will never be as fast as the simpler fixed partial charge biomolecular force fields, we are now approaching an era in which AMOEBA and similar models can be routinely used for applications that hitherto had been intractable.

ACKNOWLEDGEMENTS. We thank the Edinburgh team: Weronika Filinger, Mario Antonioletti, Lorna Smith, Arno Proeme, and Neil Chue Hong, for their help during the last 3 years in regards to early improvements in OpenMP parallelization of TINKER and issues concerning software sustainability. We also thank Jean-Phillip Piquemal for early results on the new TINKER-HP distributed memory code. The U.S. researchers thank the National Science Foundation for support of this work: AA, OND, YM, MHG and THG acknowledge grant CHE-1265731 for the software developed in TINKER in this project. THG acknowledges CHE-1363320 for the MBE work and the iEL/SCF method. MHG acknowledges CHE-1363342 for the EDA work. MET and DTM acknowledge CHE-1265889 for software developed in TINKER. DC and JS acknowledge CHE-1265704, and JWP acknowledges grant CHE-1265712 for support of work on TINKER. JWP also acknowledges National Institutes of Health grant R01-GM106137 for funding of the SAMPL4 results. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The U.K. investigators thank the Engineering and Physical Sciences Research Council (EPSRC) UK: JWE, C-KS, JD and RTB (EP/K039156/1), C-KS and JD (EP/J015059/1) and HB and IT (EP/K040529/1). JD and CKS would like to acknowledge the UKCP consortium (EPSRC grant number EP/K013556/1) for access to the ARCHER national supercomputer, and for access to the IRIDIS High Performance Computing Facility of the University of Southampton. JD acknowledges the use of supercomputing resources at the TASK Computer Centre in Gdansk, Poland.

REFERENCES

1. Lifson, S.; Warshel, A., Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and N-Alkane Molecules. *J. Chem. Phys.* **1968,** *49*, 5116-5129.
2. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.
3. Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009,** *30*, 1545-1614.
4. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005,** *26*, 1781-1802.
5. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D., et al., OpenMM 4 : A Reusable , Extensible , Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013,** *9*, 461-469.
6. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald : An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
7. Sagui, C.; Pedersen, L. G.; Darden, T. A., Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations. *J. Chem. Phys.* **2004,** *120*, 73-87.
8. Leimkuhler, B.; Reich, S., *Simulating Hamiltonian Dynamics*. Cambridge University Press: 2005.
9. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nosé–Hoover Chains: The Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **1992,** *97*, 2635-2643.
10. Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L., Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996,** *87*, 1117-1157.
11. Remsing, R. C.; Baer, M. D.; Schenter, G. K.; Mundy, C. J.; Weeks, J. D., The Role of Broken Symmetry in Solvation of a Spherical Cavity in Classical and Quantum Water Models. *J. Phys. Chem. Lett.* **2014,** *5*, 2767-2774.
12. Jungwirth, P.; Winter, B., Ions at Aqueous Interfaces: From Water Surface to Hydrated Proteins. *Annu. Rev. Phys. Chem.* **2008,** *59*, 343-366.
13. Vazdar, M.; Pluharova, E.; Mason, P. E.; Vacha, R.; Jungwirth, P., Ions at Hydrophobic Aqueous Interfaces: Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2012,** *3*, 2087-2091.
14. Fried, S. D.; Wang, L. P.; Boxer, S. G.; Ren, P.; Pande, V. S., Calculations of the Electric Fields in Liquid Solutions. *J. Phys. Chem. B* **2013,** *117*, 16236-16248.
15. Thompson, M. A.; Schenter, G. K., Excited-States of the Bacteriochlorophyll-B Dimer of Rhodopseudomonas-Viridis - a QM/MM Study of the Photosynthetic Reaction-Center That Includes MM Polarization. *J. Phys. Chem.* **1995,** *99*, 6374-6386.
16. Nerenberg, P.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T., Optimizing Solute-Water Van Der Waals Interactions to Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012,** *116*, 4524-4534.
17. Nerenberg, P. S.; Head-Gordon, T., Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011,** *7*, 1220-1230.
18. Johnson, M.; Malardier-Jugroot, C.; Murarka, R.; Head-Gordon, T., Hydration Water Dynamics near Biological Interfaces. *J. Phys. Chem. B* **2009,** *113*, 4082-4092.
19. Johnson, M. E.; Malardier-Jugroot, C.; Head-Gordon, T., Effects of Co-Solvents on Peptide Hydration Water Structure and Dynamics. *Phys. Chem. Chem. Phys.* **2010,** *12*, 393-405.
20. Demerdash, O.; Yap, E. H.; Head-Gordon, T., Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu. Rev. Phys. Chem.* **2014,** *65*, 149-174.
21. Warshel, A.; Kuwajima, S., Incorporating Electric Polarizabilities in Water-Water Interaction Potentials. *J. Phys. Chem.* **1990,** *94*, 460-466.

22. Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D., An Effective Fragment Method for Modeling Solvent Effects in Quantum Mechanical Calculations. *J. Chem. Phys.* **1996,** *105*, 1968-1986.
23. Gao, J. L., Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations. *J. Phys. Chem. B* **1997,** *101*, 657-663.
24. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A., Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from Ab Initio Quantum Chemistry. *J. Phys. Chem. B* **1999,** *103*, 4730-4737.
25. Cieplak, P.; Caldwell, J.; Kollman, P., Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases. *J. Comput. Chem.* **2001,** *22*, 1048-1057.
26. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A., Development of a Polarizable Force Field for Proteins Via Ab Initio Quantum Chemistry: First Generation Model and Gas Phase Tests. *J. Comput. Chem.* **2002,** *23*, 1515-1531.
27. Ren, P. Y.; Ponder, J. W., Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. *J. Comput. Chem.* **2002,** *23*, 1497-1506.
28. Lamoureux, G.; MacKerell, A. D.; Roux, B., A Simple Polarizable Model of Water Based on Classical Drude Oscillators. *J. Chem. Phys.* **2003,** *119*, 5185-5197.
29. Piquemal, J. P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C., Inclusion of the Ligand Field Contribution in a Polarizable Molecular Mechanics: SIBFA-LF. *J. Comput. Chem.* **2003,** *24*, 1963-1970.
30. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A., Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from Ab Initio Quantum Chemistry. *J. Phys. Chem. A* **2004,** *108*, 621-627.
31. Patel, S.; Brooks, C. L., CHARMM Fluctuating Charge Force Field for Proteins: I Parameterization and Application to Bulk Organic Liquid Simulations. *J. Comput. Chem.* **2004,** *25*, 1-15.
32. Donchev, A. G.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, O. V.; Tarasov, V. I., A Quantum Mechanical Polarizable Force Field for Biomolecular Interactions. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 7829-7834.
33. Harder, E.; Kim, B. C.; Friesner, R. A.; Berne, B. J., Efficient Simulation Method for Polarizable Protein Force Fields: Application to the Simulation of BPTI in Liquid. *J. Chem. Theory Comput.* **2005,** *1*, 169-180.
34. Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H. X.; Cieplak, P.; Duan, Y., Strike a Balance: Optimization of Backbone Torsion Parameters of Amber Polarizable Force Field for Simulations of Proteins and Peptides. *J. Comput. Chem.* **2006,** *27*, 781-790.
35. Fanourgakis, G. S.; Xantheas, S. S., The Bend Angle of Water in Iced Ih and Liquid Water: The Significance of Implementing the Nonlinear Monomer Dipole Moment Surface in Classical Interaction Potentials. *J. Chem. Phys.* **2006,** *124*, 174504.
36. Benjamin, K. M.; Schultz, A. J.; Kofke, D. A., Virial Coefficients of Polarizable Water Applications to Thermodynamic Properties and Molecular Clustering. *J. Phys. Chem. C* **2007,** *111*, 16021-16027.
37. Rasmussen, T. D.; Ren, P. Y.; Ponder, J. W.; Jensen, F., Force Field Modeling of Conformational Energies: Importance of Multipole Moments and Intramolecular Polarization. *Int. J. Quantum Chem.* **2007,** *107*, 1390-1395.
38. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P., Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramoecular Interactions and Ligand-Macromolecule Complexes. A Bottom-up Strategy. *J. Chem. Theory Comput.* **2007,** *3*, 1960-1986.
39. Geerke, D. P.; van Gunsteren, W. F., Calculation of the Free Energy of Polarization: Quantifying the Effect of Explicitly Treating Electronic Polarization on the Transferability of Force-Field Parameters. *J. Phys. Chem. B* **2007,** *11*, 6425-6436.

40. Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N., Polarization Effects for Hydrogen-Bonded Complexes of Substituted Phenols with Water and Chloride Ion. *J. Chem. Theory Comput.* **2007,** *3*, 1987-1992.
41. Salanne, M.; Vuilleumier, R.; Madden, P. A.; Simon, C.; Turq, P.; Guillot, B., Polarizabilities of Individual Molecules and Ions in Liquids from First Principles. *J. Phys. Condens. Matter* **2008,** *20*, 494207.
42. Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M., Polarization Effects in Molecular Mechanical Force Fields. *J. Phys. Condens. Matter* **2009,** *21*, 333102.
43. Patel, S.; Davis, J. E.; Bauer, B. A., Exploring Ion Permeation Energetics in Gramicidin A Using Polarizable Charge Equilibration Force Fields. *J. Am. Chem. Soc.* **2009,** *131*, 13890-13891.
44. Tafipolsky, M.; Engels, B., Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011,** *7*, 1791-1803.
45. Wang, L.-P.; Chen, J.; van Voorhis, T., Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013,** *9*, 452-460.
46. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013,** *9*, 5430-5449.
47. Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J. P., Polarizable Force Fields for Biomolecular Modeling. *Rev. Comp. Chem.* **2015,** *28*, 51-86.
48. Arismendi-Arrieta, D. J.; Riera, M.; Bajaj, P.; Prosmiti, R.; Paesani, F., I-Ttm Model for Ab Initio-Based Ion–Water Interaction Potentials. 1. Halide–Water Potential Energy Functions. *J. Phys. Chem. B* **2016,** *120*, 1822-1832.
49. Kohagen, M.; Lepsik, M.; Jungwirth, P., Calcium Binding to Calmodulin by Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2014,** *5*, 3964-3969.
50. Kohagen, M.; Mason, P. E.; Jungwirth, P., Accounting for Electronic Polarization Effects in Aqueous Sodium Chloride Via Molecular Dynamics Aided by Neutron Scattering. *J. Phys. Chem. B* **2016,** *120*, 1454-1460.
51. Freitag, M. A.; Gordon, M. S.; Jansen, J. H.; Stevens, W. J., Evaluation of Charge Penetration between Distributed Multipolar Expansions. *J. Chem. Phys.* **2000,** *112*, 7300-7306.
52. Narth, C.; Lagardere, L.; Polack, E.; Gresh, N.; Wang, Q.; Bell, D. R.; Rackers, J. A.; Ponder, J. W.; Ren, P. Y.; Piquemal, J.-P., Scalable Improvement of Spme Multipolar Electrostatics in Anisotropic Polarizable Molecular Mechanics Using a General Short-Range Penetration Correction up to Quadrupoles. *J. Comput. Chem.* **2016,** *37*, 494-505.
53. Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P., General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015,** *11*, 2609-2618.
54. Lee, A. J.; Rick, S. W., The Effects of Charge Transfer on the Properties of Liquid Water. *J. Chem. Phys.* **2011,** *134*, 184507.
55. Soniat, M.; Hartman, L.; Rick, S. W., Charge Transfer Models of Zinc and Magnesium in Water. *J. Chem. Theory Comput.* **2015,** *11*, 1658-1667.
56. Soniat, M.; Kumar, R.; Rick, S. W., Hydrated Proton and Hydroxide Charge Transfer at the Liquid/Vapor Interface of Water. *J. Chem. Phys.* **2015,** *143*, 044702.
57. Soniat, M.; Pool, G.; Franklin, L.; Rick, S. W., Ion Association in Aqueous Solution. *Fluid Phase Equilibr.* **2016,** *407*, 31-38.
58. Soniat, M.; Rick, S. W., The Effects of Charge Transfer on the Aqueous Solvation of Ions. *J. Chem. Phys.* **2012,** *137*, 044511.
59. Soniat, M.; Rick, S. W., Charge Transfer Effects of Ions at the Liquid Water/Vapor Interface. *J. Chem. Phys.* **2014,** *140*, 184703.
60. Ren, P. Y.; Ponder, J. W., Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003,** *107*, 5933-5947.
61. Ren, P. Y.; Wu, C. J.; Ponder, J. W., Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011,** *7*, 3143-3161.

62. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr., et al., Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010,** *114*, 2549-2564.
63. Young, D., *Iterative Solutions of Large Linear Systems.* Academic Press: New York, 1971.
64. Ponder, J. W. *Tinker--Software Tools for Molecular Design*, 7.0; 2014.
65. Schnieders, M. J.; Fenn, T. D.; Pande, V. S., Polarizable Atomic Multipole X-Ray Refinement: Particle Mesh Ewald Electrostatics for Macromolecular Crystals. *J. Chem. Theory Comput.* **2011,** *7*, 1141-1156.
66. Tuckerman, M.; Berne, B. J.; Martyna, G. J., Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992,** *97*, 1990-2001.
67. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015,** *11*, 1864-1874.
68. Wang, L. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S., Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013,** *117*, 9956-9972.
69. Demerdash, O. N.; Head-Gordon, T., Convergence of the Many-Body Expansion for Energy and Forces for Classical Polarizable Models in the Condensed Phase. *J. Chem. Theory Comput.* **2016,** *in press,* DOI: 10.1021/acs.jctc.6b00335.
70. Skylaris, C. K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C., Introducing ONETEP: Linear-Scaling Density Functional Simulations on Parallel Computers. *J. Chem. Phys.* **2005,** *122*, 84119.
71. Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X., et al., Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015,** *113*, 184-215.
72. Albaugh, A.; Demerdash, O.; Head-Gordon, T., An Efficient and Stable Hybrid Extended Lagrangian/Self-Consistent Field Scheme for Solving Classical Mutual Induction. *J. Chem. Phys.* **2015,** *143*, 174104.
73. Margul, D. T.; Tuckerman, M. E., A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *J. Chem. Theory Comput.* **2016,** *12*, 2170-2180.
74. Boateng, H. A.; Todorov, I. T., Arbitrary Order Permanent Cartesian Multipolar Electrostatic Interactions. *J. Chem. Phys.* **2015,** *142*, 034117.
75. Mao, Y.; Horn, P. R.; Mardirossian, N.; Skylaris, C.-K.; Head-Gordon, T.; Head-Gordon, M., Approaching the Basis Set Limit for DFT Calculations Using an Environment-Adapted Minimal Basis with Perturbation Theory: Formulation, Proof of Concept, and a Pilot Implementation. *J. Chem. Phys.* **2016,** 145, 044109.
76. Demerdash, O. N.; Mao, Y.; Head-Gordon, M.; Head-Gordon, T., J. Chem. Theory Comput. **2016** submitted.
77. Hankins, D.; Moskowitz, J., Water Molecule Interactions. *J. Chem. Phys.* **1970,** *53*, 4544-4554.
78. Stoll, H.; Preuss, H., Direct Calculation of Localized HF Orbitals in Molecule Clusters, Layers and Solids. *Theor. Chim. Acta.* **1977,** *46*, 11-21.
79. Leverentz, H.; Maerzke, K.; Keasler, S.; Siepmann, J.; Truhlar, D., Electrostatically Embedded Many-Body Method for Dipole Moments, Partial Atomic Charges, and Charge Transfer. *Phys. Chem. Chem. Phys.* **2012,** *14*, 7669-7678.
80. Richard, R.; Herbert, J., A Generalized Many-Body Expansion and a Unified View of Fragment-Based Methods in Electronic Structure Theory. *J. Chem. Phys.* **2012,** *137*, 064113.
81. Gordon, M.; Fedorov, D.; Pruitt, S.; Slipchenko, L., Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2011,** *112*, 632-672.
82. Fedorov, D. G.; Kitaura, K., Theoretical Background of the Fragment Molecular Orbital (FMO) Method and Its Implementation in GAMESS. In *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, Fedorov, D. G.; Kitaura, K., Eds. CRC: Boca Raton, FL, 2009; pp 5-36.

83. Wang, L.-P. simtk.org: Forcebalance: Systematic Force Field Optimization. https://simtk.org/home/forcebalance/, (accessed Feb. 23, 2013).
84. Wang, L.-P.; Martinez, T. J.; Pande, V. S., Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014,** *5*, 1885-1891.
85. Wang, L. P.; Chen, J. H.; Van Voorhis, T., Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013,** *9*, 452-460.
86. Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W., Revised Parameters for the Amoeba Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015,** *119*, 9423-9437.
87. Skylaris, C.-K.; Mostofi, A. A.; Haynes, P. D.; Diéguez, O.; Payne, M. C., Nonorthogonal Generalized Wannier Function Pseudopotential Plane-Wave Method. *Phys. Rev. B* **2002,** *66*, 035119.
88. Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A., Biomolecular Electrostatics and Solvation: A Computational Perspective. *Q. Rev. Biophys.* **2012,** *45*, 427-491.
89. Dziedzic, J.; Mao, Y.; Shao, Y.; Ponder, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K., To Be Submitted. **2016**.
90. Stone, A. J., Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005,** *1*, 1128-1132.
91. Vitale, V.; Dziedzic, J.; Dubois, S. M.; Fangohr, H.; Skylaris, C. K., Anharmonic Infrared Spectroscopy through the Fourier Transform of Time Correlation Function Formalism in ONETEP. *J. Chem. Theory Comput.* **2015,** *11*, 3321-3332.
92. Lever, G.; Cole, D. J.; Hine, N. D.; Haynes, P. D.; Payne, M. C., Electrostatic Considerations Affecting the Calculated HOMO-LUMO Gap in Protein Molecules. *J. Phys. Condens. Matter* **2013,** *25*, 152101.
93. Kaliman, I. A.; Slipchenko, L. V., Libefp: A New Parallel Implementation of the Effective Fragment Potential Method as a Portable Software Library. *J. Comput. Chem.* **2013,** *34*, 2284-2292.
94. Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V., Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012,** *112*, 632-672.
95. Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Flick, J.; Kaliman, I.; Shao, Y.; Gilbert, A. T.; Krylov, A. I.; Slipchenko, L. V., Effective Fragment Potential Method in Q-Chem: A Guide for Users and Developers. *J. Comput. Chem.* **2013,** *34*, 1060-1070.
96. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.
97. Simmonett, A. C.; Pickard, F. C., IV; Schaefer, H. F., III; Brooks, B. R., An Efficient Algorithm for Multipole Energies and Derivatives Based on Spherical Harmonics and Extensions to Particle Mesh Ewald. *J. Chem. Phys.* **2014,** *140*, 184101.
98. Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L., et al., PSI4: An Open-Source Ab Initio Electronic Structure Program. *WIREs Comput. Mol. Sci.* **2012,** *2*, 556-565.
99. van Dam, H. J. J.; de Jong, W. A.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M., NWChem: Scalable Parallel Computational Chemistry. *WIREs Comput. Mol. Sci.* **2011,** *1*, 888-894.
100. Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R., Optimization of Quantum Mechanical Molecular Mechanical Partitioning Schemes: Gaussian Delocalization of Molecular Mechanical Charges and the Double Link Atom Method. *J. Chem. Phys.* **2002,** *117*, 10534.
101. Wang, W.; Skeel, R. D., Fast Evaluation of Polarizable Forces. *J. Chem. Phys.* **2005,** *123*, 164107.
102. Pulay, P., Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration. *Chem. Phys. Lett.* **1980,** *73*, 393-398.
103. Belle, D. V.; Froeyen, M.; G.Lippens; Wodak, S. J., Molecular Dynamics Simulation of Polarizable Water by an Extended Lagrangian Method. *Mol. Phys.* **1992,** *77*, 239-255.
104. Lamoureux, G.; Roux, B., Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003,** *119*, 3025-3039.

105. Lopes, P. E.; Roux, B.; Mackerell, A. D., Jr., Molecular Modeling and Dynamics Studies with Explicit Inclusion of Electronic Polarizability. Theory and Applications. *Theor. Chem. Acc.* **2009,** *124*, 11-28.
106. Simmonett, A. C.; Pickard, F. C.; Shao, Y.; Cheatham, T. E., III; Brooks, B. R., Efficient Treatment of Induced Dipoles. *J. Chem. Phys.* **2015,** *143*, 074115.
107. Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, É.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P., Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *J. Chem. Theory Comput.* **2014,** *10*, 1638-1651.
108. Lagardère, L.; Lipparini, F.; Polack, É.; Stamm, B.; Cancès, É.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P., Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: II. Toward Massively Parallel Computations Using Smooth Particle Mesh Ewald. *J. Chem. Theory Comput.* **2015,** *11*, 2589-2599.
109. Niklasson, A. M., Extended Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett.* **2008,** *100*, 123004.
110. Niklasson, A. M.; Cawkwell, M. J., Generalized Extended Lagrangian Born-Oppenheimer Molecular Dynamics. *J. Chem. Phys.* **2014,** *141*, 164123.
111. Niklasson, A. M.; Steneteg, P.; Odell, A.; Bock, N.; Challacombe, M.; Tymczak, C.; Holmström, E.; Zheng, G.; Weber, V., Extended Lagrangian Born–Oppenheimer Molecular Dynamics with Dissipation. *J. Chem. Phys.* **2009,** *130*, 214109.
112. Niklasson, A. M.; Tymczak, C.; Challacombe, M., Time-Reversible Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett.* **2006,** *97*, 123001.
113. Niklasson, A. M.; Tymczak, C.; Challacombe, M., Time-Reversible Ab Initio Molecular Dynamics. *J. Chem. Phys.* **2007,** *126*, 144103.
114. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R., A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982,** *76*, 637-649.
115. Berendsen, H. J.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J., Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.
116. Tuckerman, M. E.; Parrinello, M., Integrating the Car-Parrinello Equations .1. Basic Integration Techniques. *J. Chem. Phys.* **1994,** *101*, 1302-1315.
117. Blochl, P. E.; Parrinello, M., Adiabaticity in 1st-Principles Molecular-Dynamics. *Phys. Rev. B* **1992,** *45*, 9413-9416.
118. Minary, P.; Tuckerman, M. E.; Martyna, G. J., Long Time Molecular Dynamics for Enhanced Conformational Sampling in Biomolecular Systems. *Phys. Rev. Lett.* **2004,** *93*, 150201.
119. Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E., Stochastic, Resonance-Free Multiple Time-Step Algorithm for Molecular Dynamics with Very Large Time Steps. *Mol. Phys.* **2013,** *111*, 3579-3594.
120. Leslie, M., DL_MULTI - a Molecular Dynamics Program to Use Distributed Multipole Electrostatic Models to Simulate the Dynamics of Organic Crystals. *Mol. Phys.* **2008,** *106*, 1567-1578.
121. Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T., DL_POLY_3: New Dimensions in Molecular Dynamics Simulations Via Massive Parallelism. *J. Mater. Chem.* **2006,** *16*, 1911-1918.
122. Nymand, T. M.; Linse, P., Ewald Summation and Reaction Field Methods for Potentials with Atomic Charges, Dipoles, and Polarizabilities. *J. Chem. Phys.* **2000,** *112*, 6152-6160.
123. Brommer, P.; Beck, P.; Chatzopoulos, A.; Gaehler, F.; Roth, J.; Trebin, H.-R., Direct Wolf Summation of a Polarizable Force Field for Silica. *J. Chem. Phys.* **2010,** *132*.
124. Smith, W.; Bush, I. J.; Todorov, I. T., A DAFT DL_POLY 1 Distributed Memory Adaptation of the Smoothed Particle Mesh Ewald Method. *Comput. Phys. Commun.* **2006,** *175*, 323-329.
125. Bush, I. J., The Daresbury Advanced Fourier Transform. Daresbury Laboratory 1999.
126. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1997,** *77*, 3865-3868.
127. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010,** *132,* 154104.

128. Mardirossian, N.; Head-Gordon, M., Omega B97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys.* **2014,** *16*, 9904-9924.
129. Mardirossian, N.; Head-Gordon, M., Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97M-V. *J. Chem. Phys.* **2015,** *142,* 074111.
130. Rappoport, D.; Furche, F., Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *J. Chem. Phys.* **2010,** *133*, 134105.
131. Lee, H.-S.; Tuckerman, M. E., Dynamical Properties of Liquid Water from Ab Initio Molecular Dynamics Performed in the Complete Basis Set Limit. *J. Chem. Phys.* **2007,** *126*, 164501.
132. Lee, H.-S.; Tuckerman, M. E., Structure of Liquid Water at Ambient Temperature from Ab Initio Molecular Dynamics Performed in the Complete Basis Set Limit. *J. Chem. Phys.* **2006,** *125*, 154507.
133. Ma, Z.; Zhang, Y.; Tuckerman, M. E., Ab Initio Molecular Dynamics Study of Water at Constant Pressure Using Converged Basis Sets and Empirical Dispersion Corrections. *J. Chem. Phys.* **2012,** *137*, 044506.
134. Ma, Z.; Tuckerman, M. E., On the Connection between Proton Transport, Structural Diffusion, and Reorientation of the Hydrated Hydroxide Ion as a Function of Temperature. *Chem. Phys. Lett.* **2011,** *511*, 177-182.
135. Ma, Z.; Anick, D.; Tuckerman, M. E., Ab Initio Molecular Dynamics Study of the Aqueous HOO- Ion. *J. Phys. Chem. B* **2014,** *118*, 7937-7945.
136. Berkelbach, T. C.; Lee, H.-S.; Tuckerman, M. E., Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study. *Phys. Rev. Lett.* **2009,** *103*, 238302.
137. Lee, H. S.; Tuckerman, M. E., Ab Initio Molecular Dynamics with Discrete Variable Representation Basis Sets: Techniques and Application to Liquid Water. *J. Phys. Chem. A* **2006,** *110*, 5549-5560.
138. Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R., Auxiliary Basis-Sets to Approximate Coulomb Potentials. *Chem. Phys. Lett.* **1995,** *240*, 283-289.
139. Fruchtl, H. A.; Kendall, R. A.; Harrison, R. J.; Dyall, K. G., An Implementation of RI-SCF on Parallel Computers. *Int. J. Quantum Chem.* **1997,** *64*, 63-69.
140. Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M., Fast, Accurate Evaluation of Exact Exchange: The occ-RI-K Algorithm. *J. Chem. Phys.* **2015,** *143*, 024113.
141. Manzer, S. F.; Epifanovsky, E.; Head-Gordon, M., Efficient Implementation of the Pair Atomic Resolution of the Identity Approximation for Exact Exchange for Hybrid and Range-Separated Density Functionals. *J. Chem. Theory Comput.* **2015,** *11*, 518-527.
142. Steele, R. P.; DiStasio, R. A., Jr.; Shao, Y.; Kong, J.; Head-Gordon, M., Dual-Basis Second-Order Moller-Plesset Perturbation Theory: A Reduced-Cost Reference for Correlation Calculations. *J. Chem. Phys.* **2006,** *125*, 074108.
143. Steele, R. P.; DiStasio, R. A.; Head-Gordon, M., Non-Covalent Interactions with Dual-Basis Methods: Pairings for Augmented Basis Sets. *J. Chem. Theory Comput.* **2009,** *5*, 1560-1572.
144. Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K., The SAMPL4 Host-Guest Blind Prediction Challenge: An Overview. *J. of Comput. Aided Mol. Des.* **2014,** *28*, 305-317.
145. Hamelberg, D.; McCammon, J. A., Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004,** *126*, 7683-7689.
146. Bennett, C. H., Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976,** *22*, 245-268.
147. Bell, D. R.; Qi, R.; Jing, Z.; Xiang, J. Y.; Mejias, C.; Schnieders, M. J.; Ponder, J. W.; Ren, P., Calculating Binding Free Energies of Host-Guest Systems Using the AMOEBA Polarizable Force Field. *Phys. Chem. Chem. Phys.* **2016,** *in press*, DOI: 10.1039/C6CP02509A.
148. Bradshaw, R. T.; Essex, J. W., Evaluating Parameterization Protocols for Hydration Free Energy Calculations with the AMOEBA Polarizable Force Field. *J. Chem. Theory Comput.* **2016,** *in press*, DOI: 10.1021/acs.jctc.6b00276

149. Morokuma, K., Why Do Molecules Interact - Origin of Electron Donor-Acceptor Complexes, Hydrogen-Bonding, and Proton Affinity. *Acc. Chem. Res.* **1977,** *10*, 294-300.
150. von Hopffgarten, M.; Frenking, G., Energy Decomposition Analysis. *WIREs Comput. Mol. Sci.* **2012,** *2*, 43-62.
151. Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K., Energy Decomposition Analysis Approaches and Their Evaluation on Prototypical Protein-Drug Interaction Patterns. *Chem. Soc. Rev.* **2015,** *44*, 3177-3211.
152. Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation-Theory Approach to Intermolecular Potential-Energy Surfaces of Van-Der-Waals Complexes. *Chem. Rev.* **1994,** *94*, 1887-1930.
153. Hohenstein, E. G.; Sherrill, C. D., Wavefunction Methods for Noncovalent Interactions. *WIREs Comput. Mol. Sci.* **2012,** *2*, 304-326.
154. Szalewicz, K., Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Comput. Mol. Sci.* **2012,** *2*, 254-272.
155. Kitaura, K.; Morokuma, K., A New Energy Decomposition Scheme for Molecular Interactions within the Hartree-Fock Approximation. *Int. J. Quantum Chem.* **1976,** *10*, 325-340.
156. Bickelhaupt, F. M.; Baerends, E. J., Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry. *Rev. Comput. Chem.* **2000**; *15*, 1-86.
157. Mo, Y. R.; Gao, J. L.; Peyerimhoff, S. D., Energy Decomposition Analysis of Intermolecular Interactions Using a Block-Localized Wave Function Approach. *J. Chem. Phys.* **2000,** *112*, 5530-5538.
158. Mo, Y.; Song, L.; Lin, Y., Block-Localized Wavefunction (BLW) Method at the Density Functional Theory (DFT) Level. *J. Phys. Chem. A* **2007,** *111*, 8291-8301.
159. Mo, Y.; Bao, P.; Gao, J., Energy Decomposition Analysis Based on a Block-Localized Wavefunction and Multistate Density Functional Theory. *Phys. Chem. Chem. Phys.* **2011,** *13*, 6760-6775.
160. Li, H.; Su, P., Energy Decomposition Analysis of Bonding and Nonbonding Interactions. *Abstr. Pap. Am. Chem. S.* **2009,** *238*.
161. Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M., Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A* **2007,** *111*, 8753-8765.
162. Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M., Analysis of Charge Transfer Effects in Molecular Complexes Based on Absolutely Localized Molecular Orbitals. *J. Chem. Phys.* **2008,** *128*, 184112.
163. Horn, P. R.; Sundstrom, E. J.; Baker, T. A.; Head-Gordon, M., Unrestricted Absolutely Localized Molecular Orbitals for Energy Decomposition Analysis: Theory and Applications to Intermolecular Interactions Involving Radicals. *J. Chem. Phys.* **2013,** *138*, 134119.
164. Horn, P. R.; Head-Gordon, M., Polarization Contributions to Intermolecular Interactions Revisited with Fragment Electric-Field Response Functions. *J. Chem. Phys.* **2015,** *143*, 114111.
165. Horn, P. R.; Head-Gordon, M., Alternative Definitions of the Frozen Energy in Energy Decomposition Analysis of Density Functional Theory Calculations. *J. Chem. Phys.* **2016,** *144*, 084118.
166. Horn, P. R.; Mao, Y.; Head-Gordon, M., Defining the Contributions of Permanent Electrostatics, Pauli Repulsion, and Dispersion in Density Functional Theory Calculations of Intermolecular Interaction Energies. *J. Chem. Phys.* **2016,** *144*, 114107.
167. Camilloni, C.; Sahakyan, A. B.; Holliday, M. J.; Isern, N. G.; Zhang, F.; Eisenmesser, E. Z.; Vendruscolo, M., Cyclophilin A Catalyzes Proline Isomerization by an Electrostatic Handle Mechanism. *Proc. Natl. Acad. Sci. USA* **2014,** *111*, 10203-10208.
168. Fanghanel, J.; Fischer, G., Insights into the Catalytic Mechanism of Peptidyl Prolyl Cis/Trans Isomerases. *Front. Biosci.* **2004,** *9*, 3453-3478.
169. Rezac, J.; Hobza, P., Describing Noncovalent Interactions Beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013,** *9*, 2151-2155.
170. Jurecka, P.; Sponer, J.; Cerny, 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.

171. Marshall, M. S.; Burns, L. A.; Sherrill, C. D., Basis Set Convergence of the Coupled-Cluster Correction, Delta(CCSD(T))(MP2): Best Practices for Benchmarking Non-Covalent Interactions and the Attendant Revision of the S22, NBC10, HBC6, and HSG Databases. *J. Chem. Phys.* **2011,** *135*, 194102.
172. Rezac, J.; Hobza, P., Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012,** *8*, 141-151.
173. Lao, K. U.; Herbert, J. M., An Improved Treatment of Empirical Dispersion and a Many-Body Energy Decomposition Scheme for the Explicit Polarization Plus Symmetry-Adapted Perturbation Theory (XSAPT) Method. *J. Chem. Phys.* **2013,** *139*, 034107.
174. Lao, K. U.; Herbert, J. M., Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The XSAPT Family of Methods. *J. Phys. Chem. A* **2015,** *119*, 235-252.
175. Horn, P. R.; Mao, Y., Head-Gordon, M. Probing Non-Covalent Interactions With A Second Generation Energy Decomposition Analysis Using Absolutely Localized Molecular Orbitals, *Phys. Chem. Chem. Phys.* **2016**, in press; DOI: 10.1039/c6cp03784d

TABLES

Table 1. *AMOEBA properties as a function of mutual induction convergence for a standard PCG with predictor and the iEL/SCF method using a Nosé-Hoover chain for auxiliary pseudo temperature control.*[72] Average potential energy and molecular dipole were calculated from NVT simulations at 298.0 K. Diffusion coefficients were averaged over multiple NVE simulation using independent snapshots from 298.0 K NVT simulations as the initial condition. The simulation protocol consisted of a Velocity Verlet integrator with a time step of 1.0 fs, a 9.0 Å vdW cutoff with smoothing, and a 7.0 Å real space cutoff for particle-mesh Ewald (PME) electrostatics.

Standard SCF			
Convergence (RMS Change Debye)	**Average Potential Energy (kcal/mol)**	**Average Molecular Dipole (Debye)**	**Diffusion Coefficient (10^{-5} cm^2/s)**
10^{-6}	-8.84±0.09	2.742±0.014	2.22±0.29
10^{-5}	-8.83±0.08	2.742±0.012	2.26±0.14
10^{-4}	-8.84±0.08	2.744±0.013	3.45±0.28
10^{-3}	-8.83±0.09	2.743±0.013	2.71±0.22
10^{-2}	-8.84±0.09	2.743±0.013	0.0019±0.00021
10^{-1}	-8.67±0.09	2.703±0.013	0.0020±0.00025
iEL/SCF			
Convergence (RMS Change Debye)	**Average Potential Energy (kcal/mol)**	**Average Molecular Dipole (Debye)**	**Diffusion Coefficient (10^{-5} cm^2/s)**
10^{-6}	-8.83±0.09	2.742±0.013	2.36±0.14
10^{-5}	-8.83±0.09	2.743±0.013	2.30±0.14
10^{-4}	-8.84±0.08	2.743±0.013	2.27±0.15
10^{-3}	-8.84±0.09	2.743±0.013	2.43±0.18
10^{-2}	-8.83±0.08	2.742±0.013	2.37±0.20
10^{-1}	-8.84±0.08	2.744±0.013	2.17±0.11

Table 2: *Diffusion constants for each of the stochastic isokinetic schemes described in Figure 2.* Time steps and other simulation details are given in the caption to Figure 3. For iEL/SCF the simulation protocol consisted of a Velocity Verlet integrator with a time step of 1.0 fs. All 6 reported simulations comprised a 10.4 Å vdW cutoff with smoothing, and a 10.4 Å real space cutoff for particle-mesh Ewald (PME) electrostatics (which is different than what is reported in Table 1).

Method	**Diffusion Coefficient [10^{-5} cm^2/s]**
NVE	1.9
iEL/SCF (10^{-1} D) ; NVE	2.0
NVT	2.0
XO-SIN(R)	1.8
XM-SIN(R)	1.7
XI-SIN(R)	1.8

Table 3. *CPU Performance of TINKER7 MD Simulations for iAMOEBA.* CPU timings are given for water boxes ranging from 4,800-864,000 atoms for the direct polarization model iAMOEBA. The simulations used a velocity Verlet integrator [114] with a 1.0 fs time step, a 9.0 Å vdW cutoff, and a 7.0 Å real space cutoff for particle-mesh Ewald (PME) electrostatics. Timings are reported in nanoseconds/day and speedups (in parentheses) are calculated with respect to timings corresponding to the best iAMOEBA OpenMP implementation in TINKER7. We compare the hybrid MPI/OpenMP implementation in TINKER7 using small and large numbers of cores. All timing results were obtained on a Cray XC30 using 12-core Intel "Ivy Bridge" processor at 2.4 GHz, 24 cores per node.

	Timings in ns/day for iAMOEBA		
Water System # atoms	**Standard TINKER (OpenMP only)**	**Tensor iAMOEBA (OpenMP/MPI using 384-768 cores)**	**Tensor iAMOEBA (OpenMP/MPI using 3072 cores)**
4800	2.116		8.388 (4.0)
21000	0.524	2.44 (4.7)	3.180 (6.1)
96000	0.106	0.45 (4.3)	0.760 (7.2)
288000	0.030	0.13 (4.5)	0.236 (7.9)
864000	0.008		0.062 (7.8)

Table 4. *CPU Performance of TINKER7 MD Simulations for AMOEBA and 3-AMOEBA Models.* CPU timings are given for water boxes ranging from 4,800-864,000 atoms for the mutual polarization model AMOEBA and fragments of different sized k-means clusters under the many-body approximations for 3-AMOEBA. The simulations used a velocity Verlet integrator[114] with a 1.0 fs time step, a 9.0 Å vdW cutoff, and a 7.0 Å real space cutoff for particle-mesh Ewald (PME) electrostatics. Timings are reported in nanoseconds/day and speedups (in parentheses) are calculated with respect to timings corresponding to the best AMOEBA OpenMP implementation in TINKER7. Timing results were obtained on a Cray XC30 using 12-core Intel "Ivy Bridge" processor at 2.4 GHz, 24 cores per node. All 3-AMOEBA results were fixed at 3600 cores. Adapted with permission from [69]; copyright 2016 American Chemical Society.

	Timings in ns/day for AMOEBA			
N=Number of water molecules	**Standard AMOEBA in TINKER**	**3-AMOEBA (Body=N/10 molecules)**	**3-AMOEBA (Body=N/20 molecules)**	**3-AMOEBA (Body=N/30 molecules)**
1600	1.63	5.54 (3.4)	8.04 (4.9)	6.28 (3.9)
7000	0.36	1.04 (2.9)	1.41 (3.9)	2.56 (7.1)
32000	0.07	0.23 (3.2)	0.44 (6.1)	0.54 (7.5)
96000	0.02	0.05 (2.5)	0.12 (5.9)	0.18 (9.2)
288000	0.005	0.01 (1.8)	0.04 (7.7)	0.06 (10.8)

Table 5: *GPU Performance of TINKER-OpenMM MD Simulations.* GPU timings are reported on various small to large water boxes ranging from 4,800-864,000 atoms and 3 different protein systems for the mutual polarization model AMOEBA. The simulation protocol consisted of a standard RESPA[66] integrator based on the separation of intra- and intermolecular forces with respective time steps of 1.0 fs and 2.0 fs, a 9.0 Å vdW cutoff, and a 7.0 Å real space cutoff for particle-mesh Ewald (PME) electrostatics. The single precision simulations used the TINKER-OpenMM interface on an NVIDIA GTX 980 GPU hosted on a MacPro 1.1 desktop machine. The software environment consisted of Mac OSX 10.11.4 and CUDA 7.5 and the Intel C++ and Fortran V15.0 compilers. The double precision simulations were run directly with OpenMM on an NVIDIA K80 GPU. The environment consisted of Red Hat Enterprise Linux 6.7, CUDA 7.5, and GCC 4.4.7.

		Timings in ns/day for AMOEBA		
System[a]	**# Atoms**	**OpenMM (double precision) using standard SCF solver**	**OpenMM (single precision) using standard SCF solver**	**OpenMM (single precision) using new SCF solvers**
Water (Small)	648		32.47	65.64
Crambin Crystal	1920		24.02	49.65
CBClip Complex	6432		12.45	23.22
DHFR (JAC)	23558		5.49	9.97
Water (Medium)	96000	0.201	1.458	2.29
Water (Large)	288000	0.065	0.393	0.640
Water (X-Large)	864000	0.021	0.126	0.183

[a] Details of systems: Water (Drop): 216 AMOEBA water molecules in a 18.643 Å cubic periodic box; Crambin Crystal: Two crambin chains, EtOH and water in monoclinic unit cell (PDB:3NIR); CBClip-Guest 2: Cucurbituril-based Host plus Guest 2 from SAMPL5 in 40.0 Å cubic box; DHFR (JAC): Joint Amber-CHARMM benchmark, DHFR and water in 62.23 Å cubic box; Water (Puddle): 32000 AMOEBA water molecules in a 98.65 Å cubic periodic box; Water (Lake): 96000 AMOEBA water molecules in a 142.27 Å cubic periodic box; Water (Ocean): 288000 AMOEBA water molecules in a 205.19 Å cubic periodic box.

Table 6. Mean environmental field strengths felt by the C & N atoms of the Gly-Pro peptide bond preceding the proline residue, projected along the vector perpendicular to the plane of the proline ring. AMOEBA field strengths calculated as per the protocol of Fried and coworkers.[13] The 'environment field' was taken as the difference between the fields exerted in the complex and in the substrate in the gas phase. DFT/B3LYP mean field strengths taken from Camilloni *et al.*[14]

Enzyme	**Substrate**	**Orientation**	**Field strength /MV cm^{-1}**	
			AMOEBA	**DFT**
CypA	1	*Trans*	-21.7 ± 0.2	~ -45
CypA	1	*Cis*	-16.8 ± 0.2	~ -40
CypA	2	*Trans*	-24.8 ± 0.3	~ -45
CypA	2	*Cis*	-15.8 ± 0.2	~ -40
CypB	3	*Trans*	-7.1 ± 0.2	~ -19
CypA R55A	1	*Trans*	-4.3 ± 0.2	~ -15
CypA R55A	1	*Cis*	-8.0 ± 0.2	~ -20

Peptide substrate sequences and references: 1 = HAGPIA motif from PDB entry 1M9C, 2 = GSFGPDLRAGD from reference 14, 3 = Ace-APA-Nme motiif from PDB entry 1VAI

FIGURE CAPTIONS

Figure 1: *Convergence of solute-solvent interaction energy for a single diphenylhydramine solute with 330 surrounding water molecules with reference to a DFT calculation.* The arrows indicate the number of DFT water molecules sufficient to obtain the interaction energy to within 1 kcal/mol under different embedding schemes.

Figure 2: *Electron density analysis of QM/MM using LibEFP implementation of AMOEBA.* Contour plots for the difference between the electron density of the QM water (the proton donor under the equilibrium water dimer configuration, indicated by the top figure) under the polarizing effect of the AMOEBA water, and that polarized by the Coulomb potential generated from the converged electron density (and nuclear charges) of the environement water ("Coulomb embedding"). Left: using unmodified AMOEBA water; right: using Gaussian-blurred AMOEBA water (with Gaussian exponents of O: 0.66 a_0^{-2}, H: 0.95 a_0^{-2}). The blurring is only applied to the permanent monopoles that correspond to valence electrons. Both the electron density of the QM region and the Coulomb potential of the environment water molecule are computed at the B3LYP/6-311++G(2df, 2pd) level of theory, and mutual polarization is incorporated in both QM/AMOEBA and QM/"Comloub embedding" calculations. The contours are evenly spaced at 0.1 $e^-/Å^3$, with positive contours/regions indicated by solid lines/warm colors, and negative contours/regions by dashed lines/cold colors. The black dots indicate the positions of nuclei in the QM region.

Figure 3: *Three resonance-free stochastic isokinetic Nose-Hoover (RESPA) or SIN(R) calculations are compared to an NVT benchmark* using Nose'-Hoover chains for a cubic, periodic box of length 24.85 angstroms containing 512 water molecules. Each SIN(R) calculation decomposes the forces into bonded, short-range nonbonded (including short-range induced dipoles), and long-range nonbonded terms. In the XO-SIN(R) calculation, the Nose'-Hoover coupling is applied on the long-range time scale, in the XM-SIN(R) calculation, it is applied on the short-range time scale, and in the XI-SIN(R) calculation, it is applied on the bonded time scale. In all calculations, the bonded time step is 0.5 fs, and the short-range time step is 3.0 fs. The long-range time step is 75 fs in XO-SIN(R), and it is 120 fs in the XM-SIN(R) and XI-SIN(R) calculations. In all SIN(R) calculations, the short-range electrostatic real-space cutoff is 5.0 angstroms, and the short-range van der Waals cutoff is 7.0 Å. All production runs are 150 ps long. (a) The oxygen-oxygen and (b) oxygen hydrogen radial distribution functions for all simulations using a 10.4 Å real-space cutoff. (c) the speedup obtained in each SIN(R) simulation compared to the NVT benchmark for different values of the long-range real-space cutoff. Adapted with permission from [73]; copyright 2016 American Chemical Society.

Figure 4: *RMS errors obtained from conventional SCF calculations and the MAB approach using the B97M-V density functional in a large def-QZVPPD basis, for five non-covalent datasets.* The A24 data set contains 24 small molecule intermolecular interactions energies[169], S22 contains 22 diverse intermolecular interactions[170-171], HB15 contains binding energies of hydrogen-bonded dimers featuring ionic groups common in biomolecules[172], H2O6Bind8 contains binding energies of 8 isomers of the water hexamer[173-174], and FmH2O10 contains binding energies of 10 configurations of $F^-(H_2O)_{10}$[173-174]. Adapted with permission from [75]; copyright 2016 American Institute of Physics.

Figure 5. *Calculated vs. experimental binding free energies for 14 organic ammonium guest species complexed with cucurbit[7]uril as the host.* The overall system and the numbering of the guests is the same is used for the SAMPL4 exercise.[144] The shaded region includes data points within 1.5 kcal/mol of the experimental free energy value.

Figure 6. *Distance dependence of the total interaction energy and its breakdowns (in kJ/mol) for the potassium ion-water dimer*. (a) total interaction energy; (b) permanent and induced electrostatics; (c) vdW interaction. In (a) the inset plot shows the zoomed-in near-equilibrium region in the units of *kT*, and the arrows indicate the location of energy minima for QM and AMOEBA interactions. For (b) and (c) the dash-dotted lines indicate the position of QM minimum.

Figure 7: *Root-mean-square deviations from the room-temperature crystal structure for 3 proteins simulated in explicit water.* (a) GB3 (2IGD), (b) hen lysozyme (4LZT), and (c) cyclophilin A (4YUL).

Figure 8. *Peptide substrate of cyclophilin A (cyan) in its protein environment (gray).* The environmental field, suggested to be largely due to the presence of the arginine 55 residue oriented vertically above the substrate, is perpendicular to the plane of the substrate Gly-Pro peptide bond.

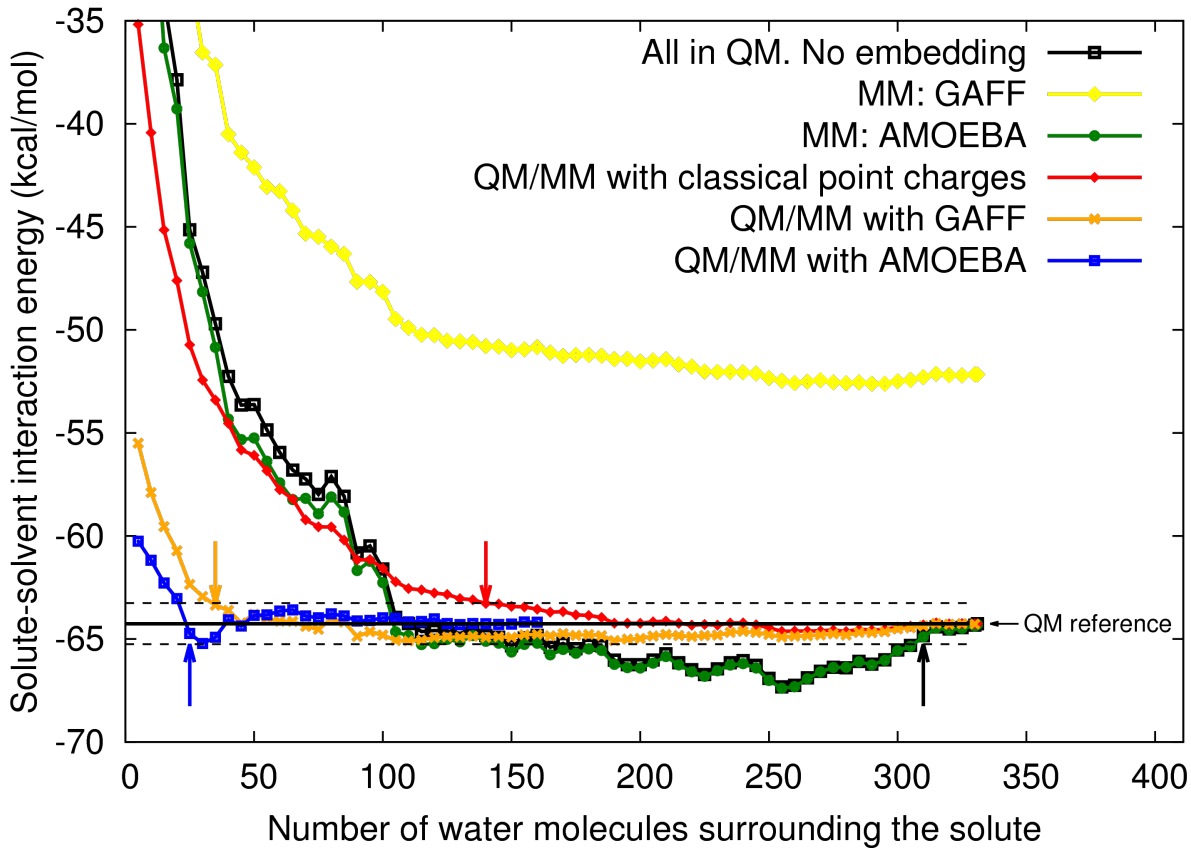


Figure 1. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

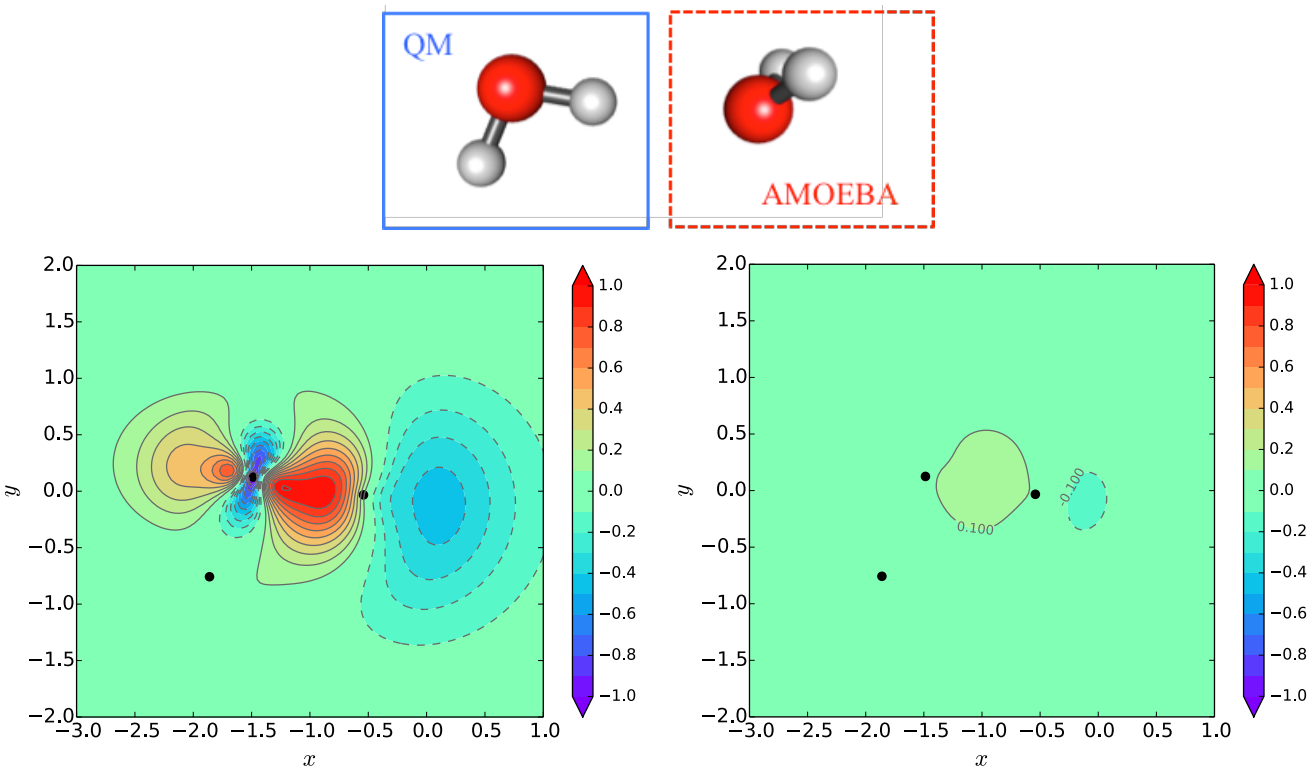


Figure 2. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

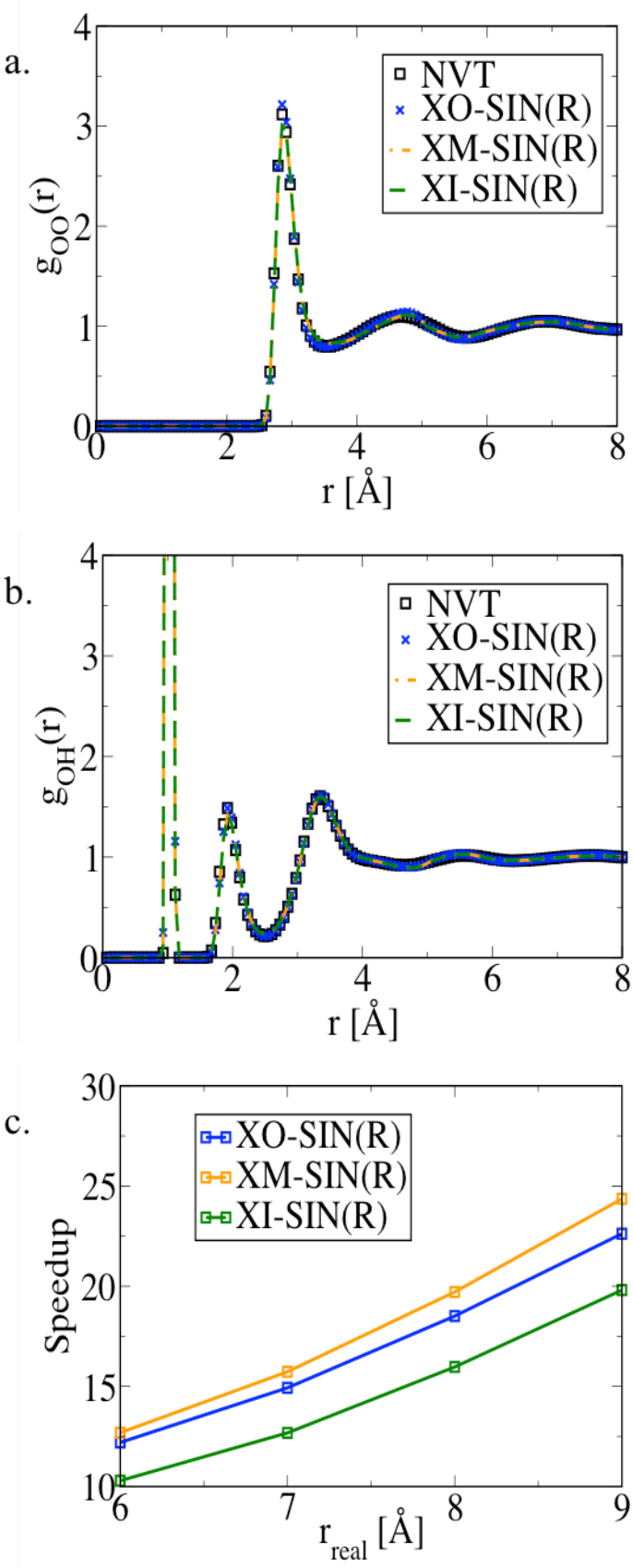


Figure 3. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

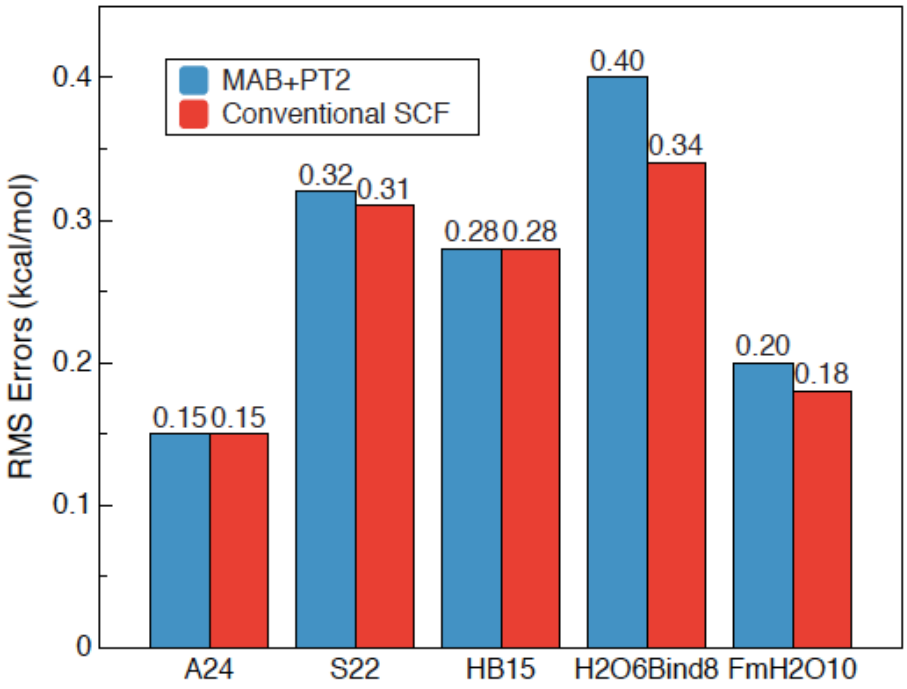


Figure 4. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

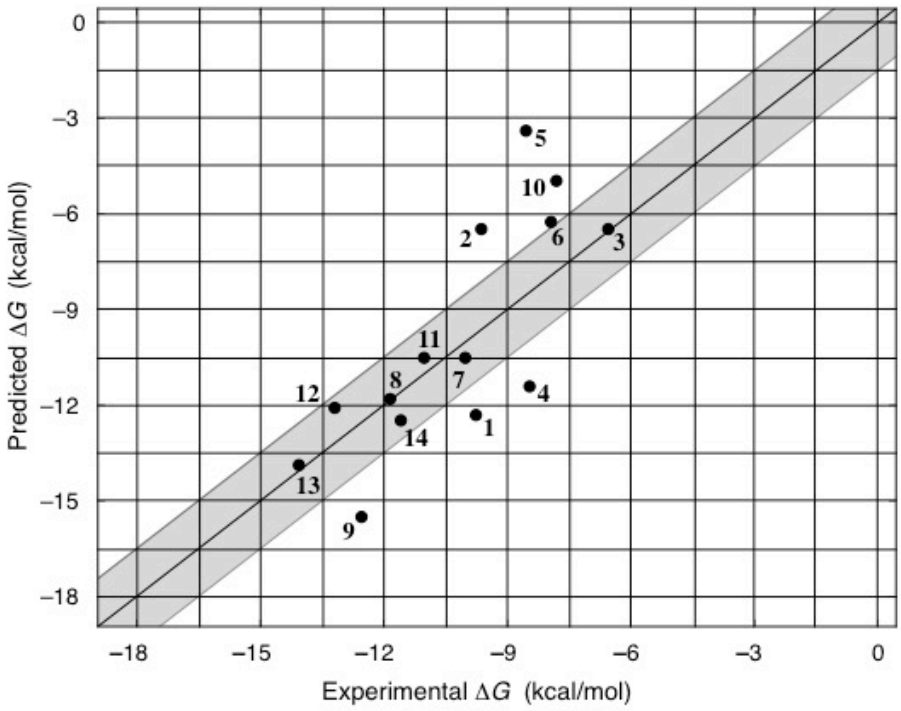


Figure 5. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

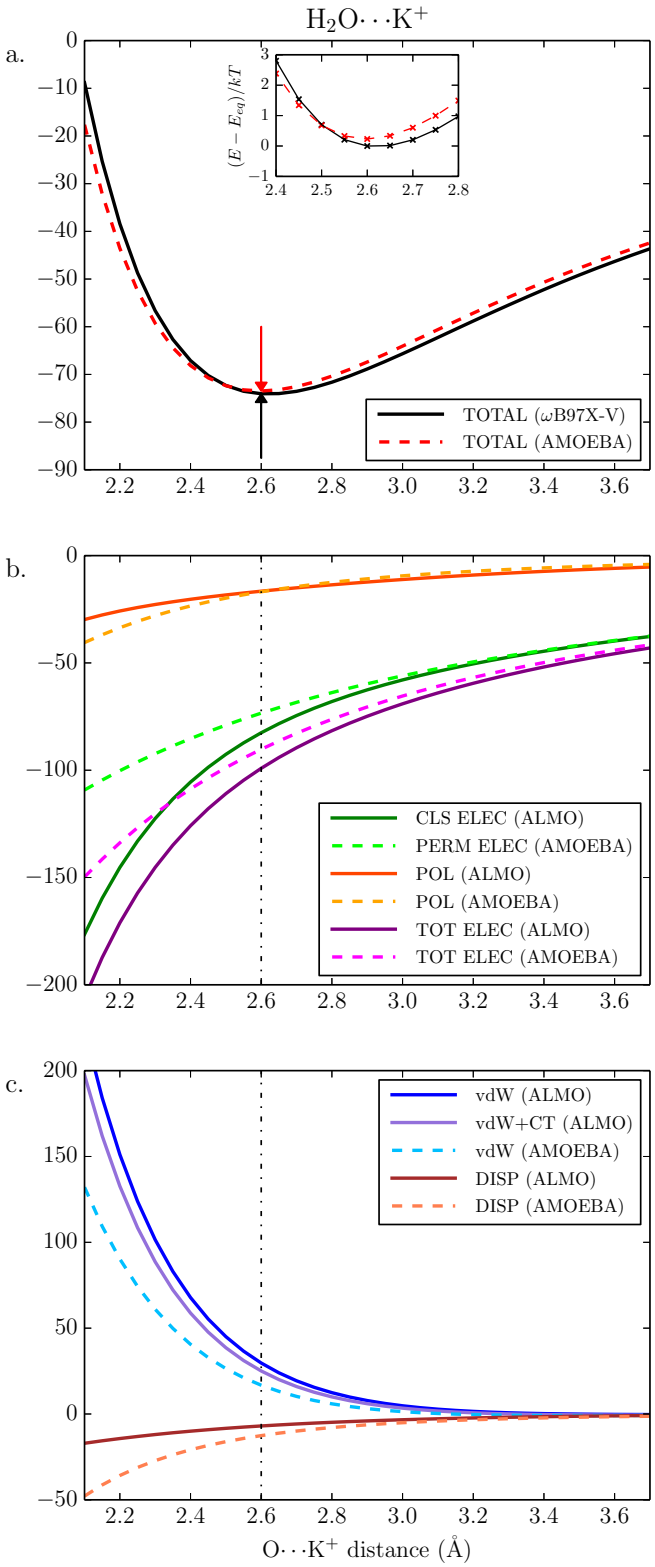
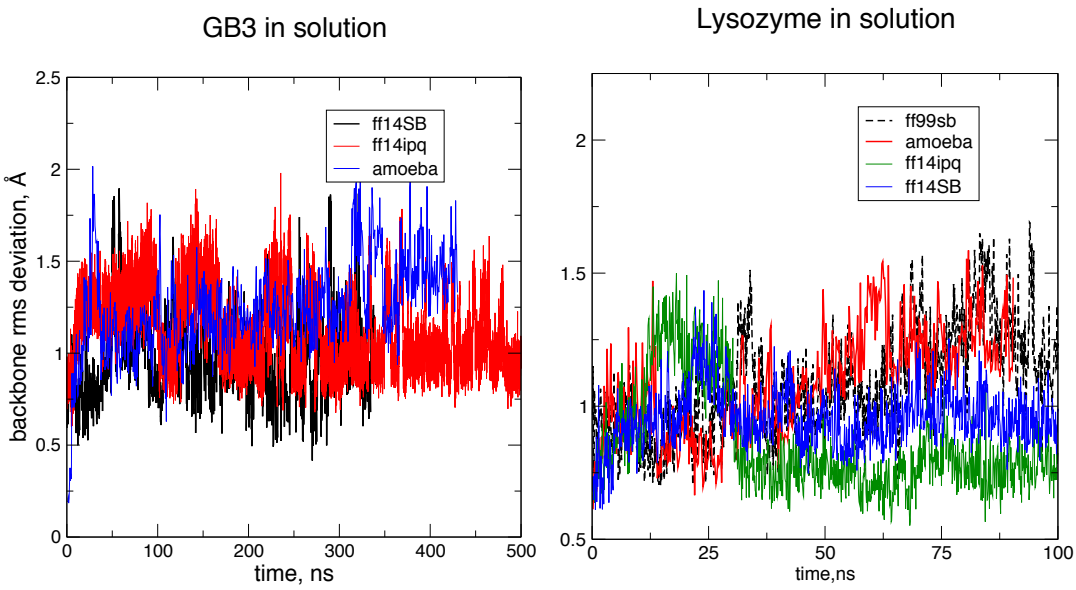


Figure 6. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

(a) (b)



(c)

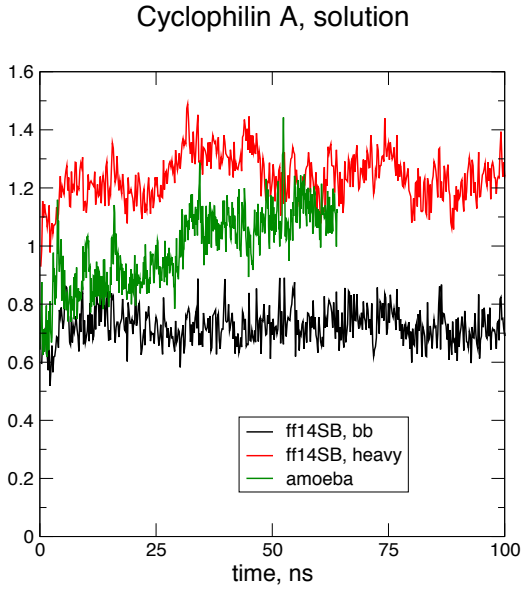


Figure 7. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers

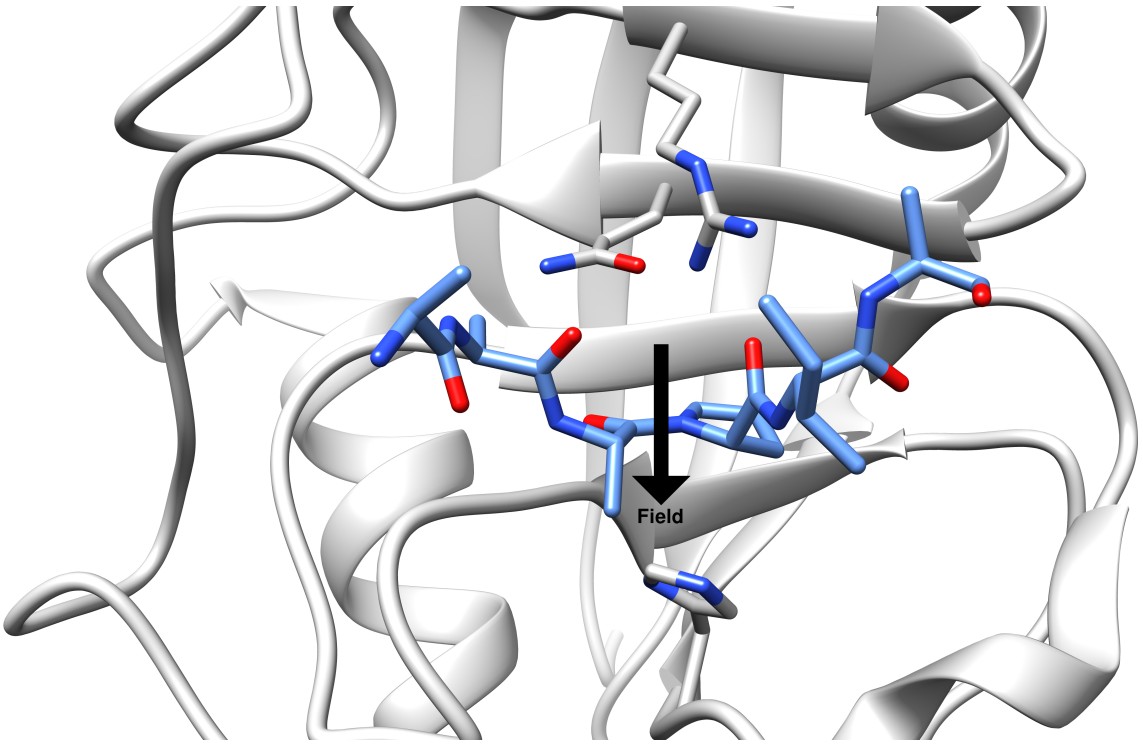
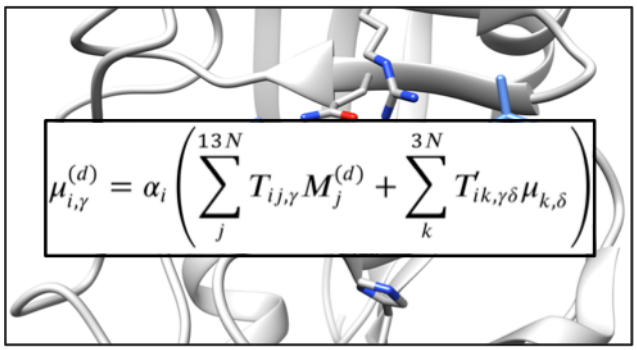


Figure 8. Albaugh, Bradshaw, Demerdash, Dziedzic, Mao, Margul and co-workers



TOC graphic

Biographies

Alex Albaugh is a graduate student in chemical engineering at the University of California, Berkeley and a member of Teresa Head-Gordon's research group. His work focuses on developing efficient methods for polarization in classical molecular dynamics simulations. He was born and raised in Palmyra, PA and is a 2012 graduate of the University of Michigan.

Henry A. Boateng is an Assistant Professor of Mathematics at Bates College in Lewiston, Maine. He received his Ph.D in Applied and Interdisciplinary Mathematics from the University of Michigan where he also served as a post-doctoral fellow. He was a computational scientist at the STFC Daresbury Laboratory, UK and a visiting professor at the University of Akron. His research interests are in method development for molecular simulations and Monte Carlo studies of crystal growth.

Richard Bradshaw is a Postdoctoral Research Fellow in Computational Systems Chemistry at the University of Southampton. Before taking up his postdoctoral role he obtained his PhD at the Institute of Chemical Biology, Imperial College London, working on joint experimental and theoretical probes of binding at protein-protein interfaces. His research interests lie in the application of advanced potentials to diverse biomolecular systems, and in the use of experimental data as a target or bias in equilibrium and enhanced sampling simulations.

Omar Demerdash is a postdoctoral scholar in chemistry in the laboratory of Teresa Head-Gordon at the University of California, Berkeley. He received his Ph.D. in Biophysics at the University of Wisconsin-Madison in 2010 under the direction of Julie Mitchell where he developed a method for predicting allosterically important residues in proteins and developed scoring functions and coarse-grained normal mode methods for protein-protein docking. He received his M.D. at the University of Wisconsin-Madison in 2013. He is currently interested in developing and applying advanced molecular mechanics potentials to problems of biological and clinical significance.

Jacek Dziedzic is a postdoctoral research associate in the School of Chemistry at the University of Southampton, and an assistant professor at the Gdansk University of Technology, where he obtained a Ph.D. in solid state physics in 2009. He is one of the principal developers of ONETEP. His research interests include linear-scaling density functional theory and hybrid exchange-correlation functionals, implicit solvation, QM/MM techniques, distributed multipole analysis and polarizable embedding.

Yuezhi Mao received his B.S. degree in material chemistry from Peking University in China in 2012. He is currently a graduate student (Ph.D. candidate) in Prof. Martin Head-Gordon's group at University of California Berkeley, where his research focuses on studying intermolecular interactions through various electronic structure methods. He is also a developer of the Q-Chem software package.

Daniel T. Margul is a computational scientist in the Structural Biology Initiative at the City University of New York Advanced Science Research Center, and a Ph.D. candidate in the Department of Chemistry at New York University. His research areas have included magnetic semiconductor synthesis, algorithms for efficient molecular dynamics simulations, and the structure and activity of proteins and nucleic acids.

Jason Swails was a postdoctoral research associate at Rutgers University after receiving his Ph.D. in Chemistry from the University of Florida in 2013. As a developer of both the Amber and OpenMM projects, he made numerous contributions in the form of expanded functionality as well as developing new methods for simulating biomolecules. He is currently working as an application developer at Lutron Electronics.

Qiao Zeng is now a visiting fellow at National Institute of Health, in the Laboratory of Computational Biology, National Heart, Lung, and Blood Institute. He obtained a Ph.D. in theoretical and computational chemistry from Xiamen University. His research interests include the development of computational methodologies in the quantum mechanical/molecular mechanical model in the field of chemistry and biochemistry.

David A. Case is in the Department of Chemistry and Chemical Biology at Rutgers University. He leads the development team for the Amber suite of biomolecular simulation programs, and has made contributions to understanding the connections between biomolecular structure and NMR and X-ray observables, to the development of implicit solvent models, and to the understanding of the electronic structure of active sites in metalloenzymes.

Peter Eastman is a Senior Software Engineer in the Stanford University Chemistry Department. He is the lead developer of OpenMM, and a contributor to several other scientific software packages. He received a Ph.D. in Applied Physics from Stanford University. He has worked in the field of scientific computing for over 20 years, especially in the areas of molecular dynamics and bioinformatics.

Jonathan Essex is Professor of Computational Chemistry at the University of Southampton. He is Chairman of the Institute for Complex Systems Simulation, and Co-Director of the EPSRC Centre for Doctoral Training in Theory and Modelling in Chemical Sciences. He obtained his D.Phil. as a Glaxo Research Scholar from the University of Oxford under the supervision of Graham Richards, before taking up a NATO Postdoctoral Fellowship with Bill Jorgensen at Yale University. His research interests like in the development and application of computer simulation methodology as applied to biological systems, with a particular interest in free energy calculations, coarse-grain and hybrid simulations, and enhanced sampling methods.

Martin Head-Gordon is the Kenneth S. Pitzer Distinguished Professor in the Department of Chemistry at the University of California, Berkeley, and a Senior Faculty Scientist in the Chemical Sciences Division of the Lawrence Berkeley National Laboratory. He arrived at Berkeley after obtaining his Ph.D. with John Pople at Carnegie-Mellon University and doing postdoctoral work with John Tully at AT&T Bell Labs. He enjoys working with his research group on the development of methods, algorithms and software in quantum chemistry, including density functional theory, wave-function based methods, and

for performing energy decomposition analysis. He also enjoys collaborative chemical applications of quantum chemistry methods in areas including catalysis, astrochemistry, and combustion related systems.

Teresa Head-Gordon is Chancellor's Professor at University of California Berkeley in the departments of Chemistry, Bioengineering, and Chemical and Biomolecular Engineering, and one of eight Co-Directors of the newly established, national NSF S2I2 Molecular Sciences Software Institute. She obtained a Ph.D. in theoretical chemistry from Carnegie Mellon University and was a postdoctoral member of technical staff at AT&T Bell Laboratories. Her research program encompasses the development of general computational methodologies applied to chemistry and biochemistry in the areas of water and aqueous hydration, chemistry at interfaces, intrinsically disordered proteins, and materials such as star polymers.

Dr. Pande is currently the Henry Dreyfus Professor of Chemistry and, by courtesy, of Structural Biology and of Computer Science at Stanford University as well as a General Partner at Andreessen Horowitz, where he runs a $200M venture fund. Prof. Pande received a BA in Physics from Princeton University in 1992 and PhD in physics from MIT in 1995. Prof. Pande's current research centers on the development and application of novel grid computing simulation techniques to address problems in chemical biology. In particular, he has pioneered novel distributed computing methodology to break fundamental barriers in the simulation of biomolecules for advances in basic biophysics as well as advances in novel therapeutics.

Jay Ponder is Professor of Chemistry, Biochemistry & Molecular Biophysics, and Biomedical Engineering at Washington University in St. Louis. He has degrees in Math and Chemistry from Wabash College, a Ph.D. in synthetic organic chemistry from Harvard University, and was a postdoctoral fellow and instructor in Biophysics at Yale University. His laboratory has a long standing interest in molecular mechanics, advances potential energy methods, conformational search algorithms, drug design and protein engineering.

Yihan Shao is a Principal Scientist at Q-Chem, Inc., and will become an Assistant Professor of Chemistry at the University of Oklahoma in August 2016. He obtained his Ph.D. in Theoretical Chemistry from the University of California at Berkeley. His research interests mainly involve the development of quantum chemistry methods and hybrid quantum mechanical molecular mechanical methods for computational chemistry/biology studies.

Chris-Kriton Skylaris is Professor of Computational Chemistry at the University of Southampton and a Fellow of the Royal Society of Chemistry. Chris obtained a first class Chemistry degree from the University of Athens and a PhD in Quantum Chemistry from the University of Cambridge. He then carried out postdoctoral research at the Physics Department in Cambridge and in 2004 he was awarded a Royal Society University Research Fellowship. His work focuses on the development of theory, algorithms and codes for quantum mechanical calculations from first principles, and their applications. He is a founding and lead author of the ONETEP program for linear-scaling quantum chemistry calculations.

Ilian Todorov leads the Computational Chemistry Group at the Scientific Computing Department, STFC Darsebury Laboratory. He and his group provide modelling and software engineering stewardship to a number of UK academic consortia - CCP5, MCC, CCPBioSim, UKCOMES - and a number of UK based industrial consortia via the Hartree Centre. Ilian obtained a Ph.D. in Computational Chemistry from University of Bristol (UK) and was a postdoctoral application lead at University of Cambridge (UK) when he started his work on the DL_POLY project. His research interests are in the area of developing sustainable software solutions for methodologies on the time and length scales of Physics, Chemistry and

Engineering. Ilian's academic interests are embodied by modelling research investigating what determines the response and change of properties of materials under irradiation, in quest of providing waste-form solutions for safe encapsulation of radioactive waste over geological timescales.

Mark Tuckerman obtained his Ph.D. in physics from Columbia University in 1993. From 1993--1994, he was an IBM postdoctoral fellow at the IBM Forschungslaboratorium in Rüschlikon, Switzerland, and from 1995--1996, he held an NSF Postdoctoral Fellowship in Advanced Scientific Computing at the University of Pennsylvania. He is currently Professor of Chemistry and Mathematics at New York University. He is the recipient of an NSF CAREER award, an NYU Whitehead Fellowship in Biomedical and Biological Sciences, an NYU Golden Dozen Award for Excellence in Teaching, the Camile Dreyfus Teacher-Scholar award, the Friedrich Wilhelm Bessel Research Award from the Alexander von Humboldt Foundation, and a Japan Society for the Promotion of Science Research Fellowship. His research interests include *ab initio* molecular dynamics studies of the transport of charge defects in hydrogen-bonded media, development of efficient path integral molecular dynamics methods, which he has applied in studies of proton and hydroxide transport and to hydrogen-storage materials, and development of enhanced sampling techniques for studying polymorphism in molecular crystals, for predicting conformational equilibria, and for mapping multidimensional free energy landscapes of complex molecules.