QuanPol: A Full Spectrum and Seamless QM/MM Program

Nandun M. Thellamurege, Dejun Si, Fengchao Cui, Hongbo Zhu, Rui Lai, and Hui Li*

The quantum chemistry polarizable force field program (QuanPol) is implemented to perform combined quantum mechanical and molecular mechanical (QM/MM) calculations with induced dipole polarizable force fields and induced surface charge continuum solvation models. The QM methods include Hartree–Fock method, density functional theory method (DFT), generalized valence bond theory method, multiconfiguration self-consistent field method, Møller–Plesset perturbation theory method, and time-dependent DFT method. The induced dipoles of the MM atoms and the induced surface charges of the continuum solvation model are self-consistently and variationally determined together with the QM wavefunction. The MM force field methods can be user specified, or a standard force field such as MMFF94, Chemistry at Harvard Molecular Mechanics (CHARMM), Assisted Model Building with Energy Refinement (AMBER), and Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA). Analytic gradients for all of these methods are implemented so geometry optimization and molecular dynamics (MD) simulation can be performed. MD free energy perturbation and umbrella sampling methods are also implemented. © 2013 Wiley Periodicals, Inc.

Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) style methods originally developed by Warshel and Levitt[1] are practical methods for studying solvent effects. Several popular force field packages, such as AMBER,[2–6] CHARMM,[7–9] and BOSS,[10] contain semiempirical QM methods. Many force field packages, such as AMBER,[2–6] CHARMM,[7–9] GROMACS,[11–14] and TINKER,[15] have established modular architecture or interface to perform QM/MM calculations with various external quantum chemistry packages such as GAUSSIAN09,[16] GAMESS-US,[17,18] and Q-CHEM.[19] Many quantum chemistry and physics packages, for example, ADF,[20] GAMESS-UK,[21] GAUSSIAN09,[16] NWCHEM,[22] and CP-PAW,[23] have integrated MM methods and QM/MM capabilities.

However, none of the QM/MM programs and interfaces mentioned above can work for a full spectrum of QM methods and MM force field methods in an additive coupling. For example, none of them is able to run molecular dynamics (MD) simulations using time-dependent density functional theory method (TDDFT[24,25]) and second-order Møller–Plesset perturbation theory (MP2[26]) methods additively coupled with general induced dipole polarizable force fields and induced surface charge implicit solvation models.

In this article, we describe the implementation and test of the quantum chemistry polarizable force field program, QuanPol, which is designed to perform QM/MM style calculations with general *ab initio* electronic structure methods and general induced dipole polarizable force field.[27–31] Using a polarizable force field for QM/MM methods (i.e., QM/MMpol methods) is advantageous because the electronic polarization of the MM region can be explicitly included.[31–43] The QuanPol program is integrated into the general atomic and molecular electronic system (GAMESS) package.[17,18] CHARMM,[7–9] AMBER,[2–6] OPLS-AA,[44,45] and MMFF94[46–49] force field potential energy functions, including induced dipole polarization and induced surface charge reaction field, are implemented. A full spectrum of standard quantum chemical methods can be used. These methods are Hartree–Fock (HF), DFT methods[50], generalized valence bond theory method (GVB[51]), multiconfiguration self-consistent-field method (MCSCF[52–55]), MP2[26] methods, and TDDFT[24,25]. Rigorous analytic energy gradients are derived and implemented for all of these QM/MM style methods. Gradient and Hessian-guided geometry optimization and various MD simulation can be performed. Virtually, all MM and QM/MM calculations are parallelized.

Theory

QM/MMpol and QM/MMpol/continuum potential energy

In a previous work, we implemented variational SCF QM/MMpol and QM/MMpol/continuum style HF, DFT, GVB, and MCSCF methods and their analytic gradients.[56] HF-based MP2 method is a relatively accurate electronic structure method for recovering electron correlation energy.[26] TDDFT method formulated by Casida et al.[24,25] is an efficient and relatively accurate QM method for studying valence and singe excited states. We previously implemented analytic gradient for QM/continuum style MP2 and TDDFT methods that incorporates induced surface charges[57,58] and QM/MMpol style MP2 and TDDFT methods that incorporate induced dipoles.[59,60] Recently, we derived and implemented analytic gradients for QM/MMpol/continuum style MP2 and TDDFT methods, the details of which will be published in forthcoming technical papers. The QuanPol program is based on these theories and methods.

N. M. Thellamurege, D. Si, F. Cui, H. Zhu, R. Lai, H. Li
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588
Email: hli4@unl.edu

In the QuanPol implementation, the total energy of a molecular system described with a QM/MMpol/continuum style HF, DFT, GVB, and MCSCF methods can be schematically written as

$$E_{QM/MMpol/C} = E_{QM} + E_{MM} + E_{chg} + E_{rep} + E_{disp} + E_{pol} + E_{sol} \quad (1)$$

E_{QM} in eq. (1) is the electronic energy (including nuclear repulsion) of the QM region calculated using the regular gas-phase Hamiltonian for these methods but with the wavefunction perturbed by the MM atoms and the continuum solvent.

E_{MM} in eq. (1) contains the MM covalent potential terms and pair-wise intermolecular interaction (nonpolarizable charge-charge and Lennard–Jones (LJ) or their variants) potential terms. We implemented in QuanPol the bond stretching, bending, dihedral angle torsion, dihedral angle bending, wagging, bond stretch-bending, as well as the CHARMM peptide dihedral crossterm energy correction map (CMAP[61]) potential functions. In the implementation, we used the efficient mathematical formulas recommended by Tuzun et al.[62]

E_{chg} in eq. (1) is the electrostatic interaction energy between QM electrons/nuclei and MM atomic point charges:

$$E_{chg} = \sum_{\mu v} P_{\mu v} V_{chg,\mu v} + E_{Z,chg}$$

$$= \sum_{\mu v} P_{\mu v} \sum_{q} \left\langle \mu(\mathbf{r}_e) \left| \frac{q}{|\mathbf{r}_e - \mathbf{r}_q|} \right| v(\mathbf{r}_e) \right\rangle + \sum_{Z} \sum_{q} \frac{Z \cdot q}{|\mathbf{r}_Z - \mathbf{r}_q|} \quad (2)$$

In eq. (2), μ and v are spin-orbital basis functions. $V_{chg,\mu v}$ is the one-electron integrals of the MM point charges q, $E_{Z,chg}$ is the electrostatic energy between QM nuclear charges Z and MM point charges q. $\mathbf{r}_Z$, $\mathbf{r}_e$, and $\mathbf{r}_q$ are the coordinates of the QM nuclei, QM electron, and MM point charges. P is the one-particle density matrix in the HF, DFT, GVB, and MCSCF methods.

E_{rep} and E_{disp} in eq. (1) are the repulsion and dispersion interaction energies, respectively, between QM atoms and MM atoms. QuanPol can use standard force field LJ potentials to model them:

$$E_{rep} + E_{disp} = E_{LJ} \quad (3)$$

We note that the LJ potential does not directly affect the QM electronic wavefunction.

Alternatively, QuanPol can use effective Gaussian potentials $a \cdot exp(-b \cdot R^2)$ (here R is the distance between the electronic coordinate and the MM atom when used to calculate the one-electron integrals, a and b are two parameters) at each MM atom to describe the short-range repulsion and some short-range (but not long-range) dispersion interactions between QM atoms and MM atoms.

$$E_{rep} + E_{disp} = \sum_{\mu v} P_{\mu v} V_{rep,\mu v} + \sum_{\mu v} P_{\mu v} V_{disp,\mu v}$$

$$= \sum_{\mu v} P_{\mu v} \sum_{rep} \left\langle \mu(\mathbf{r}_e) \left| a_{rep} \cdot exp\left(-b_{rep} \left| \mathbf{r}_e - \mathbf{r}_{rep} \right|^2 \right) \right| v(\mathbf{r}_e) \right\rangle \quad (4)$$

$$+ \sum_{\mu v} P_{\mu v} \sum_{disp} \left\langle \mu(\mathbf{r}_e) \left| a_{disp} \cdot exp\left(-b_{disp} \left| \mathbf{r}_e - \mathbf{r}_{disp} \right|^2 \right) \right| v(\mathbf{r}_e) \right\rangle$$

V_{rep} and V_{disp} are the QM one-electron integrals of the effective Gaussian repulsion potential, effective Gaussian dispersion

potential of the MM atoms. $\mathbf{r}_e$, $\mathbf{r}_{rep}$, and $\mathbf{r}_{disp}$ are the coordinates of the QM electron, MM repulsion, and dispersion points (i.e., MM atoms). The repulsion and dispersion potential have the same form, but with different a and b parameters: repulsion has positive a and smaller b, whereas dispersion has negative a and larger b. Again, here P is the one-particle density matrix in the HF, DFT, GVB, and MCSCF methods.

The effective Gaussian potential method in QuanPol is identical to that in the effective fragment potential (EFP[63]) methods for modeling QM–MM repulsion interactions. It is not related to the "Gaussian Blur"[64] method for QM–MM electrostatic interactions. Compared to LJ method, the effective Gaussian potential method has a direct impact on the QM wavefunction, thus is closer to QM description. The disadvantage is that the effective Gaussian potential cannot describe long-range dispersion interactions, and its parameters are not readily available for standard force field methods.

The MM atomic charges and effective Gaussian potentials directly perturb the QM electronic wavefunction. However, they are "static" MM potentials that do not respond to QM wavefunction changes. In QuanPol, these potentials are included in the QM calculation as one-electron integrals, and are precalculated before the SCF iterations. E_{pol} in eq. (1) is the induced dipole polarization energy of the MM atoms in the QM/MMpol or QM/MMpol/continuum system:[56]

$$E_{pol} = -\frac{1}{2} \mathbf{F}^T \mathbf{d} \quad (5)$$

Here, $\mathbf{d}$ is a vector collecting the induced dipoles of all MM atoms, $\mathbf{F}$ is a vector collecting the electric fields at all MM atoms. The superscript "T" denotes the transpose. Equation (5) is the standard energy expression used in induced dipole polarizable force fields. For QM/MMpol style HF, DFT, GVB, and MCSCF methods, the electric fields $\mathbf{F}$ have contributions from QM nuclei, QM electrons and MM atomic charges, and the electric fields due to other induced dipoles and induced surface charges must not be included. The induced dipole at each MM atom is calculated as the product of the nine-component dipole polarizability tensor of the MM atom and the total electric field (now including the electric field due to other induced dipoles and the electric field due to all induced surface charges) at the MM atom. All induced dipoles and induced surface charges are iterated to reach self-consistence. In QuanPol, only the isotropic (scalar) dipole polarizability tensors can be used so the nine-component polarizability tensor is actually a diagonal matrix with the same value for x, y, and z directions. The charge–dipole and dipole–dipole interactions are excluded if two MM atoms are connected by one or two covalent bonds (1–2 and 1–3 exclusion rule).

E_{sol} in eq. (1) is the induced surface charge solvation free energy of the QM/MMpol molecule in the QM/MMpol/continuum system:[56]

$$E_{sol} = \frac{1}{2} \mathbf{V}^T \mathbf{q} \quad (6)$$

Here, $\mathbf{q}$ is a vector collecting all induced surface charges, and $\mathbf{V}$ is a vector collecting all electric potentials at the induced

surface charges. The superscript "T" denotes the transpose. Equation (6) is the standard energy expression used in induced surface charge solvation models. The electric potentials have contributions from QM nuclei, QM electrons, and MM atomic charges. The electric potentials due to other induced surface charges and induced dipoles must not be included. In QuanPol, E_{sol} is calculated with a modified conductorlike screening model (COSMO[65] or CPCM[66,67]) program called FixSol,[68] which is implemented for general MMpol and QM/MMpol systems. Using the FIXPVA2 tessellation scheme,[68] the nuclear gradient in MMpol/FixSol and QM/MMpol/FixSol methods can be evaluated analytically. The induced surface charge at each surface point is determined by the total electric potential (now including the electric potential due to other induced surface charges and the electric potential due to all induced dipoles) at the surface point. All induced dipoles and induced surface charges are iterated to reach self-consistence.

When both induced dipoles and induced surface charges are involved in a QM/MMpol/continuum calculation, the induced dipole energy and induced surface charge energy have exactly the same expressions as in eqs. (5) and (6).[56] When the induced dipoles **d** and induced surface charges **q** are solved, the interactions between **d** and **q** must be considered. So, adding a continuum solvation model to a induced dipole force field will affect the magnitudes of all induced dipoles and their energy, and vice versa. The total $E_{pol} + E_{sol}$ is not the simple addition of the individual E_{pol} and E_{sol} calculated without each other. We have used the "super-matrix" concept to describe the induced dipoles and induced surface charges using one linear equation.[56]

In QuanPol, the induced dipoles and induced surface charges are included into HF, DFT, GVB, and MCSCF methods using a variational SCF scheme, in acoordance with the SCF nature of these QM methods. To derive these QM/MMpol/continuum SCF methods, one must first write down the total energy (incuding induced dipole energy and induced suface charge energy) as a function of the electron density matrix, then take the partial derivatives with respect to the wavefunction coefficients to obtain the generalized Fock operators. Compared to regular gas phase Fock operators, the generalized Fock operators contain the potentials of MM charges, effective Gaussian potentials, induced dipoles, and induced surface charges. As discussed above, the MM charges and effective Gaussian potentials are "static" potentials. The induced dipoles and induced surface charges are "responsive" potentials depending on the electron density. In QuanPol, they are solved iteratively and self-consistently together with the electron density in SCF calculations. After the SCF convergence, the wavefunction, induced dipoles and induced surface charges are determined, and the induced dipole polarization energy and the induced surface charge solvation free energy are determined as well. The computational procedure is:

a. Calculate the electric fields at all dipole polarizabilities and the electric potentials at all surface charge grid points. The electric fields and potentials have contributions from QM nuclei, QM electrons, and MM atomic charges.

b. Using an iterative scheme to solve the induced dipoles and induced surface charges. In general, the induced dipoles and induced surface charges can affect each other through dipole–dipole, dipole–charge, charge–charge interactions. Therefore, the electric fields at all dipoles due to other dipoles and surface charges should be calculated, so as the eletric potentials at all surface charges due to other surface charges and dipoles. We have used the "super-matrix" concept to describe the induced dipoles and induced surface charges using one linear equation.[56]

c. Calculate the one-electron integrals of the potentials of the induced dipoles and induced surface charges, and add them to the Fock operator, then solve for a new wavefunction.

d. If the wavefunction is not converged, go back to step (a). When the wavefunction is converged, the induced dipoles and induced surface charges are also converged.

Using the above variational treatment is advantageous because analytic gradient can be readily evaluated. We have derived and implemented analytic gradient for all of these QM/MMpol/continuum style HF, DFT, GVB, and MCSCF methods. It turned out that the forces involving the induced dipoles and induced surface charges can be evaluated as if they were permanent point dipoles and charges.

Using a variational scheme also enables post SCF treatments. For example, linear response (i.e., time-dependent) and perturbation theory methods within the QM/MMpol and QM/MMpol/continuum frames can be formulated. We have derived and implemented analytic gradient for QM/continuum style MP2 and TDDFT methods that incorporates induced surface charges[57,58] and QM/MMpol style MP2 and TDDFT methods that incorporate induced dipoles.[59,60] Recently, we derived and implemented analytic gradients for QM/MMpol/continuum style MP2 and TDDFT methods. Here, these two methods are briefly described. Interested readers are refered to our recent papers.[57–60]

For HF-based MP2 methods (including closed shell, spin-restricted open shell, and spin-unrestricted open shell cases), the second-order correction energy $E^{(2)}$ is obtained by using the same formulas for regular gas-phase MP2 methods but with the molecular orbitals and orbital energies from QM/MMpol/continuum style HF calculations. The total energy of the system is simply the HF/MMpol/continuum energy plus the $E^{(2)}$:

$$E_{MP2/MMpol/C} = E_{HF/MMpol/C} + E^{(2)}$$
$$= E_{HF/MMpol/C} + E^{(2)}_{ele} + E^{(2)}_{nuc} + E^{(2)}_{chg} + E^{(2)}_{rep} + E^{(2)}_{disp} + E^{(2)}_{pol} + E^{(2)}_{sol} \quad (7)$$

The second-order correction energy $E^{(2)}$ contains the corrections to the electron–electron interaction energy $E^{(2)}_{ele}$, the electron–nucleus interaction energy $E^{(2)}_{nuc}$, and the electron-MM/continuum interaction energies ($E^{(2)}_{chg}$, $E^{(2)}_{rep}$, $E^{(2)}_{disp}$, $E^{(2)}_{pol}$, and $E^{(2)}_{sol}$). The $E^{(2)}$ is actually computed as a single energy term, and is not easily separable. So the energy terms in eq. (7) are only schematic. To calculate the MP2 analytic gradient, the induced

dipoles and induced surface charges must be included into the Z-matrix method to solve for the MP2 "response density matrix."

For TDDFT, the induced dipoles and induced surface charges must be included in the calculation of the excitation energy E^ω to allow mutual responses among QM, MM, and the continuum solvent (the continuum solvent may show a fast electronic response). The total energy can be written as:

$$E_{\text{TDDFT/MMpol}/C} = E_{\text{DFT/MMpol}/C} + E^\omega$$

$$= E_{\text{DFT/MMpol}/C} + E^\omega_{\text{ele}} + E^\omega_{\text{nuc}} + E^\omega_{\text{chg}} + E^\omega_{\text{rep}} + E^\omega_{\text{disp}} + E^\omega_{\text{pol}} + E^\omega_{\text{sol}} \quad (8)$$

The excitation energy E^ω contains the contributions from electron–electron interaction energy E^ω_{ele}, the electron–nucleus interaction energy E^ω_{nuc}, and the electron-MM/continuum interaction energies (E^ω_{chg}, E^ω_{rep}, E^ω_{disp}, E^ω_{pol}, and E^ω_{sol}). The E^ω is computed as a single energy term and is not easily separable. So the energy terms in eq. (8) are only schematic. To calculate the TDDFT analytic gradient, the induced dipoles and induced surface charges must be included into the Z-matrix equations for solving the "difference density matrix" and the "transition state density matrix."

Compared to various implementations of QM/MM and QM/MM/continuum methods in the literature, the QM/MMpol/continuum method implemented in the QuanPol program is probably the most general one. For example, when induced dipole polarizable force fields are used to form QM/MMpol methods, the QM methods are usually HF and DFT, or their semiempirical variants. In QuanPol, we implemented our latest QM/MMpol style TDDFT and MP2 methods with analytic gradients, thus expanded the boarderline of QM/MM methods. When induced surface charge implicit solvation models are used to form QM/MM/continuum methods,[69–72] induced dipole polarization is usually not included, and the QM methods are usually limited to HF and DFT, or their semiempirical variants. In addition, not all QM/MM/continuum methods have analytic energy gradient implemented. The QuanPol QM/MMpol/continuum method rigorously combines induced dipole polarizable force fields and induced surface charge solvation models into various QM calculations, and has analytic gradients. It is probably the most sophisticated implementation that can find many applications.

QM–MM interactions across covalent bond

In QuanPol QM/MM calculations, QM atoms will feel the intermolecular interaction potentials (i.e., charge, induced dipole, LJ, or effective Gaussian potential) of the MM atoms. There should be no force field charge and/or dipole polarizability for QM atoms, and no LJ interactions between QM atoms. The repulsion dispersion between QM and MM atoms should be described using either LJ model [eq. (3)] or effective Gaussian potentials on the MM atoms [eq. (4)].

If there is no covalent bond between the QM and MM regions, there should be no force field covalent terms within the QM region. If there are covalent bonds between QM atoms and MM atoms, the MM covalent interactions will remain in full strengths if they involve at least one MM atom. The MM atoms that form covalent bonds with QM atoms should not have effective Gaussian potentials, which are intended to mimic the intermolecular repulsion–dispersion interactions. The QM electrons will feel the full strengths of the MM intermolecular interaction potentials (i.e., charge and induced dipoles) of the MM atoms that form covalent bonds with QM atoms. These interactions will not over polarize the QM electronic wavefunction when the following capping H atom method is used.

In QuanPol, a capping H atom method can be used to treat covalent bond cuts between QM and MM regions. For example, when a protein C_α—C_β bond is cut, an H atom is put in the place of the $C\alpha$ to saturate the open bond of the side chain (QM region). The C_α atom retains its MM covalent parameters in the QM/MM calculation, but should not retain its force field charge, polarizability, LJ potential, and/or effective Gaussian potentials. The force on the C_α atom is the sum of the QM force on the H atom and the MM force on the C_α atom. In all calculations, the H and C_α atoms are enforced to have the same coordinate, mass of carbon, force, and velocity. No additional degrees of freedom are introduced. A main difficulty in using H to replace C_α is that their QM bond lengths do not match: a normal C—H bond length is ~1.1 Å, whereas the C_α—C_β bond length is ~1.5 Å. QuanPol can use one or several effective Gaussian repulsion potentials in the form of $a \cdot \exp(-b \cdot R^2)$ for the capping H atom to produce a H—C bond length of ~1.5 Å. These potentials are similar to those used for MM atoms to mimic the intermolecular repulsion–dispersion interactions between QM and MM atoms, but here they are used for the capping QM H atoms to create repulsion and longer bond lengths. These potentials act on the electrons of all QM atoms (including the capping H atom itself) in the QM region via one-electron integrals, thus are generally good for different QM methods, basis sets, and different electronic states. Different Gaussian potentials can be used for different atoms to produce different bond lengths.

We have tested the accuracy of this scheme. For the CH_3CH_2OH and $CH_3CH_2O^-$ pair, stretching a methyl H—C bond from 1.10 to 1.51 Å with a single Gaussian potential ($a = 3.0$ e/bohr and $b = 6.0$ bohr^{-2}) on the H atom leads to very small changes in the deprotonation energy. For example, MP2 and B3LYP[73] (both with the aug-cc-pVDZ basis set[74]) give deprotonation energies of 383.19 and 383.25 kcal/mol, respectively, for the normal (and optimized) molecules, and of 384.46 and 384.10 kcal/mol, respectively, for the "H-C stretched and optimized" molecules. The errors are 1.27 and 0.85 kcal/mol, similar to those in the pseudobond scheme developed by Zhang et al.[75] and an improved pseudobond scheme developed by Zhang.[76] The gas-phase deprotonation reaction used here is probably an extreme case in QM/MM calculations. In general cases, the capping H atom and effective Gaussian potential method has errors much smaller than 1.27 kcal/mol.

To compensate for the weakening of the covalent terms due to the elongation of the bond involving the capping H atom, the bond, angle, and dihedral rotation (only these three) terms involving other QM atoms and the capping H atom (but

no MM atoms) are scaled by 0.5 or any user input value from 0 to 1. The scaling factor 0.5 is determined empirically. When a single Gaussian potential with $a = 3.0$ e/bohr and $b = 6.0$ bohr^{-2} is used for the H atom of a C—H bond, the optimized C—H bond length elongates from 1.10 to 1.50 Å and the bond stretching and bending force constants are reduced by ~ 0.4 times. We performed geometry optimization and Hessian harmonic frequency calculations for methane (CH_4) at the MP2/aug-cc-pVDZ level of theory, and obtained the following nine frequencies in cm^{-1}: 1323.71, 1323.82, 1323.87, 1550.71, 1550.75, 3063.25, 3206.55, 3207.06, and 3207.13. By adding a single-Gaussian potential with $a = 3.0$ e/bohr and $b = 6.0$ bohr^{-2} to a H atom, we obtained the following nine frequencies in cm^{-1}: 845.76, 845.77, 1133.35, 1446.40, 1446.48, 1952.31, 3073.88, 3188.40, and 3188.48. Clearly, after the bond length elongation, a C—H stretching mode frequency is reduced from ~ 3207 to ~ 1952 cm^{-1}. Because frequency is proportional to the square root of the force constant, the affected C—H bond stretching force constant is weakened by approximately $(1952/3207)^2 = 0.37$ times. The lowest two frequencies are determined by the angle bending force constants involving the elongated C—H bond, which are weakened by approximately $(846/1324)^2 = 0.41$ times after elongation. Therefore, both the bond stretching and bending force constants are weakened by ~ 0.4 times. In general, the force constants involving a H—C bond is $\sim 10\%$ larger than those involving a C—C bond. So, the bond stretching and bending force constants involving a C_α—C_β bond should be scaled down by ~ 0.5 to accommodate the QM contribution.

The above three examples (CH_3CH_2OH, $CH_3CH_2O^-$, and CH_4) are pure QM molecules with no MM atoms. In QM/MM calculations, the capping H atom will feel the charges and induced dipoles of the MM atoms, especially those close to the capping H atom. For a covalent cut at a protein C_α—C_β bond, we observed that the MM interactions tend to make the H—C bond shorter by 0.1 Å, and found that a single-Gaussian potential with $a = 3.0$ e/bohr and $b = 3.0$ bohr^{-2} at the capping H atom is required to produce the desired 1.54 Å bond length. It is not easy to know how this Gaussian potential affects the stretching, bending, and dihedral rotation force constants in such a QM/MM calculation. We therefore assume that these H—C force constants in the QM calculation are weakened by ~ 0.4 times, and the corresponding C_α—C_β force constants in the MM calculation should be scaled down by ~ 0.5 times. We implemented 0.50 as the default scaling factor in QuanPol for C—C single bonds. Users can specify their own values from the input file.

Senn and Thiel[77] reviewed various boundary schemes like "link H atom," "boundary atom," and "frozen localized orbitals." The capping H atom method implemented in QuanPol is similar to the "link H atom" scheme, but does not introduce additional degrees of freedom. It is also similar to the "boundary atom" scheme, but does not need sophisticated effective or pseudopotentials.

When covalent bonds are cut, the MM region is often left with a fraction of charge. The missing charge can be added to the MM atoms that form covalent bonds to QM atom. If many such MM atoms are involved, the missing charge can be evenly distributed to them. These MM atoms will typically receive $+0.01$ to $+0.02$ e charges in the case of a protein C_α—C_β bond cut. QuanPol can automatically handle this issue but a user can always modify the charges through the input deck. Charge shifting[78] and charge redistribution[79] methods have been developed to zero off the charge of the MM atoms that are the closest to the "link H atom" between QM and MM atoms, and to preserve the local dipole moment. The magnitudes of the MM charges involved in these methods are typically 0.1–0.5 e. These methods are necessary for avoiding the overpolarization of the QM wavefunction when there is a close contact (~ 0.5 Å) between QM and MM atoms. The QuanPol capping H atom scheme (together with the effective Gaussian potential for maintaining a longer bond) is insensitive to the charges of the nearby MM atoms (which are typically 1.54 Å away), and does not require zeroing off their charges. In addition, because the charge modifications on the MM atoms are typically 0.01–0.02 e, it is not necessary to reconstruct the local dipole moment.

Long-range interactions

Shifting and switching functions are often used to smoothly cut off long-range electrostatic and dispersion interactions. Usually shifting functions work for the full range from zero to the cutoff, while switching functions work at the end region of the cutoff. We implemented in QuanPol several options to apply shifting and switching functions to MM atom–atom interactions.

Switching functions can also be used in QM/MMpol style calculations. It is possible to use only one point to represent a relatively small QM molecule (i.e., group switching). Because the QM molecule is not rigid, it is difficult to use the center of mass or anyone of the QM atoms to represent the QM molecule. This difficulty can be resolved by using a fixed point to represent the QM molecule and let the QM molecule stay around the fixed point in the MD simulation. The details of this method can be found in a recent paper.[59] This method uses switching function for all one-electron integrals. Because the switching functions are applied directly to the one-electron integrals of the basis functions, this approach can be used for any QM methods. We use the initial geometric center of the QM atoms to represent the QM molecule, and the QM molecule must stay around its initial position in the MD simulation. This is done by applying a bowl-like potential V to one (and only one) of the atoms of the QM molecule:

$$\text{if } r \leq r_0, V = 0$$
$$\text{if } r > r_0, V = k(r - r_0)^2 \tag{9}$$

Where r is the distance of the atom to the fixed point. In QuanPol, $k = 1.0$ hartree/bohr^2 and $r_0 = 3.78$ bohr (2.00 Å) are implemented. For rigorous energy conserving MD simulations, the fixed point is always fixed. However, when volume scaling and velocity scaling are used, energy conservation is not strictly enforced, the fixed point is actually allowed to update

(in the sense of a position scaling) at every MD step. In Quan-Pol, this QM-MM style switching function is used for QM–LJ, QM–charge, and QM–dipole interactions.

Ewald summation[80] is a method for calculation of electrostatic interaction of periodic systems. We implemented in QuanPol the conductor (i.e., tin-foil) boundary condition Ewald summation for charge–charge interactions in pure MM calculations (no QM atoms). Compared to shifting and switching function methods, Ewald summation needs additional computing time for the reciprocal term. When induced dipole polarizable force field is used, Ewald summation is used only for the charge–charge interaction. Shifting or switching functions are used for the induced dipole terms (charge-dipole and dipole-dipole). We will implement modern particle-mesh Ewald summation method[81] in QuanPol for MM and QM/MM methods.

MD simulation

We implemented in QuanPol a method developed by Zhou et al.[82] to assign initial velocities to all mass points. The Verlet[83] integration algorithm and a modified Beeman integration algorithm[84] are implemented in QuanPol. The Berendsen thermostat/barostat[85] and the Andersen thermostat[86] are implemented to control the temperature and pressure of a MM or QM/MM system.

QuanPol contains subroutines to evaluate some key properties from MD simulation. Volume is meaningful only when periodic boundary condition (PBC) or spherical boundary condition (SBC) is used. When rectangular PBC is used, the volume is simply the product of the three box sizes. When SBC is used, particles feel force at the boundary so they remain in the sphere. The particle distribution density decreases to zero at the boundary. More particles can be at the boundary at higher pressure and higher temperature, thus occupying larger volume. In QuanPol, the SBC volume is printed out as the volume of the sphere because the exact volume is not easy to determine.

Pressure P is computed with the virial equation:

$$P = \frac{1}{3V}\left[\sum_i^N m_i(\mathbf{v}_i \cdot \mathbf{v}_i) - \sum_i^N \sum_{j>i}^N (\mathbf{r}_{ij} \cdot \mathbf{f}_{ij}) + U_{\text{Ewald}}^{\text{reci}}\right] \quad (10)$$

Here, V is the volume of the master box, $\mathbf{r}_{ij}$ is the vector from mass point i to point j, $\mathbf{f}_{ij}$ is the pair-wise force from mass point i to point j. In QuanPol, pressure is computed from all atomic forces, including all covalent terms (bond stretching, angle bending, dihedral rotation, dihedral bending, wagging, CHARMM CMAP), LJ term, charge-charge term, and induced dipole terms (charge-dipole and dipole-dipole). $U_{\text{Ewald}}^{\text{reci}}$ is the reciprocal term in Ewald summation. The application of eq. (10) is straightforward for nonpolarizable MM force field systems. For induced dipole polarizable MM force field systems, the forces between induced dipoles are actually pairwise, so eq. (10) can also be used.

For QM/MMpol systems, we always use a bowl-like potential at a fixed Cartesian point i [see discussion around eq. (9)] to confine the QM region, and only sum up the virial $\mathbf{r}_{ij} \cdot \mathbf{f}_{ij}$ for all

of the MM mass points j. The virial between QM atoms is ignored. This will result in some underestimation of the true pressure of the entire QM/MM system. However, as the number of QM atoms is typically 1% of the total number of atoms, this will not cause significant errors.

Some times, it is useful to fix some QM and MM atoms in their Cartesian coordinates in MD simulations. In QuanPol, up to 200 QM and/or MM atoms can be fixed. The fixed atoms will be assigned initial velocities, and their initial Cartesian coordinates and velocities will remain absolutely unchanged: internal RATTLE[86] constraining, temperature, and volume/pressure scaling will have no effects on these atoms. When QM atoms are fixed, no further action is taken because the virial between QM atoms is already ignored. When MM atoms are fixed, the virial between fixed MM atoms is ignored. This, of course, will result in some underestimation of the true pressure of the entire QM/MM system. Therefore, we do not recommend the use of too many fixed atoms. In addition, we note that when fixed atoms are used, their velocities are not zeroed off but remain at the initial values, so they still have contributions to kinetic energy, temperature, and pressure [the ideal gas term in eq. (10)]. Therefore, with fixed atoms and a slightly underestimated pressure, the MD can still be run using regular ensembles such as constant number of particle, temperature and pressure (NTP) and constant number of particle, volume and energy (NVE). If a large number of atoms are fixed and accurate pressure is still needed, the volume of the fixed atoms should be accurately estimated and removed from the total volume [V in eq. (10)], and the velocities of the fixed atoms should be zeroed off [the ideal gas term in eq. (10)].

QuanPol contains subroutines to evaluate other properties in the MD simulation, such as radial distribution function, radial density profile (for SBC only), root mean square displacement of atomic coordinates from the initial coordinates, and radius of gyration (mass-weighted root mean square distances of the atoms from the center of mass), dielectric constant, and self-diffusion constant.

The dipole moment of the QM atoms, the dipole moment of user selected MM atoms, and the dipole moment of the entire QM/MM system are printed out at every MD step. The QuanPol Vibrational Spectrum Program (QPVIB) can be used to generate the time autocorrelation functions of these dipole moments and the corresponding infrared (IR) spectra. The "Fastest Fourier Transform in the West" program (FFTW3)[88] is used in the QPVIB program to perform the discrete Fourier transformation.

Geometry optimization

Two methods are implemented in QuanPol to perform MM and QM/MM geometry optimization, including transition state search. Both methods can use the RATTLE algorithm to constrain distances in geometry optimization processes.

One is an MD-like steepest descent geometry optimization algorithm. The MD nature makes it easy to connect to MD calculations. In QuanPol, the RATTLE algorithm is combined into this algorithm to constrain distances. RATTLE affects velocity,

force (acceleration), and coordinate. At every optimization step, an auxiliary velocity is computed for each atom from the force (acceleration a) using a time step size Δt (e.g., 5 fs):

$$v = \frac{1}{2} a \Delta t \qquad (11)$$

The velocities are subject to RATTLE constraints. After the RATTLE changes, the acceleration (and force) is calculated back from the changed velocities:

$$a = \frac{2v}{\Delta t} \qquad (12)$$

This new acceleration is then used to test convergence and predict the new coordinate:

$$x' = x + \frac{1}{2} a (\Delta t)^2 \qquad (13)$$

The new coordinates are also subject to RATTLE constraints. The MD time step size Δt determines the efficiency of the optimization process. For every optimization job, the initial Δt is set to be 5 fs. At every optimization step, if the total potential energy is lower than the previous step, Δt is increased by 10%, otherwise Δt is decreased by 50%. The maximum displacement is restricted to be ± 0.20 bohr. This method can only be used for locating minima (not transition state) and is not very efficient as compared to Hessian-guided methods.

The other is a Hessian-guided optimization algorithm. To use the RATTLE algorithm to constrain general distances, at every optimization step an auxiliary velocity is computed using eq. (11) for each atom from the acceleration (force) using a time step size Δt (e.g., 1 fs). The velocities are subject to RATTLE constraints. After the RATTLE change, the acceleration (and force) is calculated back using eq. (12) from the changed velocity. The RATTLE-affected force (internal force) is then used to test convergence and then used by a Hessian-guided optimization algorithm (the quadratic approximation[89] together with the Broyden–Fletcher–Goldfarb–Shanno Hessian updating scheme) implemented in GAMESS to predict new coordinates, which are subject to RATTLE constraints, too. The auxiliary time step size Δt here does not affect the results. This Hessian method can be used for locating equilibrium geometry and transition state. We note that the RATTLE procedure here does not affect the Hessian-guided optimization process except for generally constraining some distances, for example, all bond lengths involving hydrogen atoms, or internal geometry of rigid water molecules.

Free energy perturbation simulation

We have implemented MM and QM/MM MD free energy perturbation (FEP) simulation methods in QuanPol. The FEP formula proposed by Zwanzig[90] is used.

$$\Delta G = -k_B T \ln \left\langle \exp\left(-\frac{\Delta E}{k_B T} \right) \right\rangle \qquad (14)$$

Here, k_B is Boltzmann constant, T is the bath temperature. ΔG is the Gibbs free energy change from state A to state B

(constant pressure and constant temperature), and the ΔE is the potential energy change from state A to state B, which can be calculated using pure MM methods or QM/MM methods. In general, the difference between state A and state B can be molecular geometry, electronic state, or even different types and numbers of atoms. FEP methods can be formulated to calculate free energy profile along reaction coordinates (i.e., potential of mean force, PMF), solvation free energy and binding free energy, relative reduction potential, and acid-base ionization constant.

We implemented both single topology and dual topology MD simulation methods for FEP calculations. These methods have been extensively discussed in the literature.[91,92] When single topology method is used, only one set of atoms are used to run the MD simulation. Some atoms are selected to be "perturbation atoms," and are given a second set of force field parameters (e.g., mass, charge, LJ potential, bond stretching force constant, bond angle bending force constant, and others), or the same set of force field parameters but a second set of Cartesian coordinates. At every MD step, a single point energy calculation is performed using the second set of parameters or coordinates for the "perturbation atoms" to obtain the ΔE for eq. (14). Very often FEP simulations are performed for a hypothetic "alchemical" state lying between states A and B, with a mixing coefficient λ for state B and a mixing coefficient $(1 - \lambda)$ for state A. In QuanPol, force field covalent potential parameters F (e.g., equilibrium bond lengths and bond force constants) of a hypothetic state are obtained by using eq. (15).

$$F(\lambda) = (1 - \lambda)F_A + \lambda F_B \qquad (15)$$

Force field charge-charge and LJ potential energies, as well as QM electronic energies E (including nuclear repulsion), are linearly combined by using eq. (16).

$$E(\lambda) = (1 - \lambda)E_A + \lambda E_B \qquad (16)$$

Clearly, both eqs. (15) and (16) give the correct pure state A and state B when λ is 0 and 1, respectively.

When dual topology method is used, two sets of "perturbation atoms" coexist in the MD system (but those of state A not seeing those of state B) and sample different local phase spaces. Soft-core charge-charge and LJ potentials can be used to avoid sampling difficulty arising from singularity. In QuanPol, dual topology is only implemented for pure and nonpolarizable MM methods. The covalent potential parameters (e.g., equilibrium bond lengths and bond force constants) are in full strength for the "perturbation atoms" of both state A and state B (i.e., ideal-gas-molecule end states), while charge-charge and LJ potential energies are combined with eq. (16). In a dual topology MD simulation, the two sets of "perturbation atoms" should stay roughly in the same cavity but feel slightly different interaction potential energies from the environment. We note that dual topology chaperoned alchemical QM/MM free energy simulation has been developed, for example, by Yang et al.[93] Currently, QuanPol does not have the capability of performing dual topology QM/MM FEP simulations.

Several types of free energy changes can be calculated using QuanPol. For a pure MM system, one can specify the ΔE be the difference in the charge–charge and LJ interaction energy between the "perturbation atoms" and all other atoms in the system. The resulted free energy change thus is solvation free energy change. Both single and dual topology methods can be used to calculate solvation free energy change. In addition to the solvation potential energy of the "perturbation atoms," the covalent and noncovalent potential energy within the "perturbation atoms" can also be included into ΔE. In this case, one can calculate the free energy change associated with a geometric change of the "perturbation atoms." The resulted free energy change thus is the PMF. In QuanPol, two sets of slightly different Cartesian coordinates for a group of "perturbation atoms" can be specified and fixed in the MD simulation. In the MD simulation, the "perturbation atoms" always remain fixed as initially given for state A. At every MD step, a single-point energy calculation is performed using the "perturbation atoms" in coordinates initially given for state B to obtain the ΔE for eq. (15).

For QM/MM systems, the change of the total QM/MM potential energy ΔE associated with the state change (e.g., oxidation state and protonation state, but no geometry change) of the "QM perturbation atoms" can be used, and the free energy change corresponds to "rigid geometry" reduction potential E^0 or acid-base ionization constant pK_a. If the state change is in the geometry (not oxidation or protonation state) of the "QM perturbation atoms," the resulted free energy change is PMF or "geometry relaxation" free energy change. To calculate pK_a and E^0 values, both the "rigid geometry" free energy changes and the "geometry relaxation" free energy changes are significant and should be considered.

It is worth comparing the QuanPol MD FEP implementations to other methods/programs. When pure MM methods are used, the QuanPol implementation should be identical or very similar to other MM MD FEP methods. When QM/MMpol methods are used, QuanPol performs "brute force" QM/MMpol MD FEP simulations and obtains the free energy change of the whole QM/MMpol system. When accurate DFT and MP2 methods are used, such a "brute force" method can in principle give accurate results because the QM wavefunction is optimized in the MM environment, and the QM–MM interactions are not simplified. The disadvantage is that it is slow. In some other QM/MM FEP methods, due to different purposes, various simplifications are used to increase the computational efficiency. For example, in the method developed by Reddy et al.,[94] semiempirical QM methods were used without MM to compute the energy and forces for the solute. Either semiempirical or better QM methods were used to generate atomic charges for the solute. Then, the MM solvent molecules were added, and the QM–MM interaction energy and forces were calculated using atomic charges and LJ potentials of the solute. The main advantage of these methods is that the MD simulation is fast.

Umbrella sampling

We implemented umbrella sampling methods[95,96] in QuanPol, and wrote a service program called "QuanPol Weighted Histo-gram Analysis Program" (QPWHA) to use the weighted histogram analysis method (WHAM)[97] to analyze both one-dimensional (1D) and 2D histograms and generate PMF curves.

In QuanPol, the umbrella sampling bias potential is a harmonic potential applied to a reduced or combined internal coordinate R:

$$V_{bias} = \frac{1}{2} k_{bias} (R - R_0)^2 \qquad (17)$$

Here, R_0 is the desired equilibrium value of R. For example, one can set $R_0 = 3.0$ Å for a pair of Na^+ and Cl^- ions. k_{bias} is the force constant of the harmonic potential, typically 100 kcal/mol/Å^2 for distance constraint.

In QuanPol, four types of internal coordinates can be defined for 1D umbrella sampling: the distance between two atoms, the difference between two distances (involving four or three atoms), the angle between three atoms, and the dihedral rotational angle for four bonded atoms. 2D umbrella sampling can use any combination of these four types (thus 16 types). If selected, the umbrella sampling bias potential is added to all MM and QM/MM (including the mean field QM/<MM> discussed in the next subsection) MD simulation and geometry optimization calculations as if it were a real physical potential. 1D histograms are printed out with 61 bins and bin size of either 0.01 Å or 1.0 degree. 2D histograms are printed out with 3721 bins and bin size of either 0.01 Å or 1.0 degree for each dimension. These window sizes are coded in the QuanPol program and a user cannot change them from the input deck.

Mean field QM/<MM> method

Combined QM/MM MD simulation is costly when a large number of QM calculations are required. Several mean field QM/MM methods have been developed to address this issue.[98–102] To distinguish mean field QM/MM methods from regular QM/MM methods, we denote mean field methods as QM/<MM>. The central idea of QM/<MM> is that the average electronic energy of the QM region in a period of time can be approximated as the electronic energy of the QM region in the presence of the mean MM electric field in the same period of time. In these methods, one expensive QM calculation is performed following n (typically hundreds to tens of thousands) steps of inexpensive MM calculations, drastically reducing the computational time.

We implemented an average position QM/<MM> method in the QuanPol program for a full spectrum of standard QM methods.[103] In this method, the QM atoms are fixed in their Cartesian coordinates, only the MM atoms move in the MD simulation. The QM region is given a set of charges (as well as dipoles and quadrupoles) and LJ potentials to interact with the MM atoms. We have implemented code to use either force field atomic charges or multipole expansion points from QM calculations for the QM atoms, and found that the latter is much more accurate. Except for the fixation of the QM atoms and a slight underestimation of the true pressure, the MD simulation can be performed in several standard ensembles (such as constant number of particle, temperature and pressure

(NPT)). Numerical results suggest that the average position scheme works very well for typical solute-solvent systems and protein-water systems. The average position QM/<MM> method can finally reduce the computational time to a level that is only 2–5 times higher than pure MM methods.

We have formulated QM/<MM> MD FEP methods that can be used to calculate relative reduction potential (E^0), acid ionization constant (pK_a), binding free energy, and PMF profiles. The details of these methods and their applications will be presented in forthcoming papers. In these QM/<MM> MD FEP methods, the "solvation free energy change" of the QM atoms is obtained from MM MD simulation. The QM/<MM> calculation provides two energy (enthalpy) terms, the electronic energy of the QM atoms and the electrostatic interaction energy between QM atoms and MM atoms, which are combined with the "solvation free energy change" of the QM atoms to produce the "free energy change" of the QM atoms in the QM/MM system. The average electrostatic interaction between QM and MM atoms computed at the force field level is subtracted to avoid a double counting. The dynamic contribution to the free energy of the QM atoms is obtained from QM Hessian harmonic frequency calculation together with a continuum solvation model.

The QuanPol QM/<MM> MD FEP method is very similar to the first step in Hu, Lu, and Yang's QM/MM minimum free-energy path (MFEP[104]) method: the "solvation free energy change" of the QM atoms is obtained from classic MD with the QM atoms fixed, and a QM/<MM> calculation is performed to obatin the QM energy and a better description of the QM-MM electrostatic interaction energy (enthalpy). Their method focuses on the "free energy change" as a function of QM coordinates, so they computed the "free energy gradient" to geometically optimize the QM atoms for the purpose of establishing the free energy profile along a reaction coordinate. Our method, however, focuses on the "free energy change" as a function of a general perturbation. Therefore, our method can be used to compute pK_a, E^0, and PMF. For reaction coordinate free energy profile calculation, their method is able to determine the MFEP, whereas our method depends on user defined conformations.

Implementation

General description

We implemented the QuanPol program in the GAMESS package.[16,17] In August 2013, QuanPol has more than 65,600 lines of Fortran code in 164 subroutines. Same as GAMESS, QuanPol uses keywords (currently 149 keywords) to control all calculations. All the QM methods work with QuanPol are those already implemented in GAMESS by other authors. The one-electron integral routines in the QuanPol program were adopted and modified from the GAMESS code, especially the EFP code.[63] The QuanPol QM/MMpol and QM/MMpol/FixSol style MP2 method is implemented for the serial RMP2 and UMP2 program,[105] the parallel RMP2 program implemented by Fletcher et al.,[106] the parallel RMP2 program implemented by Ishimura et al.,[107,108] the parallel ZAPT2 program implemented by Fletcher et al.,[109] and Aikens et al.,[110] and the parallel UMP2 program implemented by Aikens et al.[111,112] The QuanPol QM/MMpol and QM/MMpol/FixSol style TDDFT method is implemented for the TDDFT program implemented by Chiba et al.[113,114]

QuanPol automatically performs QM/MM calculation when the numbers of QM atoms and MM atoms are both greater than zero. QuanPol can also be used to run pure QM calculations if the MM atoms are made to be dull atoms with no interactions. Double precision is used for all real variables in all QuanPol routines for MM and QM/MM calculations. QuanPol MD jobs can be restarted if the coordinates and velocity from the trajectory file are included into the next input file.

Parallel calculation

QuanPol uses the duplicated copy method to run parallel calculations for the MM part. Each compute core stores a full copy of the atomic coordinates, velocities, forces, force field topology, and parameters. At every MD step, each compute core calculates and accumulates a portion of all the pairwise energy and forces of the whole MM system, then performs a global summation across all cores to generate the final energy and forces at each core. This requires one major network communication (for the forces) per MD step for a nonpolarizable force field method. For induced dipole polarizable force field methods, the pairwise charge–dipole and dipole–dipole interactions are usually computed iteratively (typically ~10 iterations) within one MD step to reach self-consistency. Using such an iterative polarization scheme, induced dipole force fields need ~10 major network communications at every MD step.

Using the duplicated copy method, the full set of the MM atomic coordinates/velocities on each core must be updated at each MD step using either the Verlet or Beeman integration algorithm. In principle, this can be done in parallel: each core computes a subset of the new coordinates/velocities and broadcasts its subset to all other cores. However, unless the network is extremely fast, it takes more time to broadcast the new coordinates/velocities than computing all of them individually. Therefore, in QuanPol the MD integration part is not parallelized. Instead, each core updates its own full set of MM atomic coordinates and velocities with no communication with other cores. In principle, given the same initial coordinates and velocities, and the same forces at every MD step, each core is able to follow the same trajectory with exactly the same coordinates and velocities on each core at every MD step. However, due to possible machine errors, especially when different types of CPU cores are mixed, it is likely that the coordinates and velocities on different cores will be different after many MD steps. To eliminate this potential problem, at every 200 MD steps, all cores will be given the coordinates and velocities of the master core. This only requires a marginal addition of network communication time.

Pair-wise force field terms are evenly distributed to each core. All covalent terms, such as bond stretching, angle bending, stretch-bending, dihedral rotation, dihedral bending,

wagging, and CHARMM CMAP, are parallelized like this. When calculating charge–charge, charge-induced dipole, dipole–dipole, and LJ interactions, QuanPol uses Verlet neighbor list[83] method to achieve high efficiency in both memory and speed. Each compute core generates its own neighbor list. In other words, when the neighbor list is generated, each core generates a sublist. For example, for a system with 100,000 atoms, if each atom is to include 3000 nearby atoms as its neighbors, the size of the neighbor list is $2 \times 100,000 \times 3000 \div 2 = 300,000,000$, which requires 300 megawords or 2.4 gigabits of RAM. If 32 cores are used, each core will only need 10 megawords or 0.08 gigabits of RAM. Each core needs only to calculate its own list of pairs for noncovalent interaction terms.

In most force fields, the noncovalent interactions for 1–2 and 1–3 atom pairs are excluded because they have been integrated into the bond stretching and angle bending terms. In some force fields, 1–4 pairs, that is, atom pairs already considered for dihedral rotation terms, are scaled down. QuanPol uses different scaling factors for different force fields. For CHARMM, 1–4 charge interaction is not scaled, but some 1–4 LJ interactions are given a second set of LJ parameters. QuanPol identifies these 1–4 cases and apply the second set of LJ parameters. For AMBER, all 1–4 charge–charge interactions are scaled by 0.833333 (i.e., divided by 1.2), and all 1–4 LJ interactions are scaled by 0.5. For OPLS-AA, all 1–4 charge–charge and LJ interactions are scaled by 0.5. For MMFF94, all 1–4 charge–charge interactions are scaled by 0.75, but all 1–4 LJ (note MMFF94 uses 14–7 LJ potential) interactions is not scaled. Therefore, 1–2, 1–3, and 1–4 exclusion lists must be generated. In a straightforward implementation, these exclusion lists are subtracted from the neighbor list so the neighbor list does not contain any 1–2, 1–3, or 1–4 pairs. In QuanPol, such a "subtraction process" is not used. Instead, the proper noncovalent terms are obtained by subtracting the "full terms" computed using the full neighbor list (containing all of the unwanted 1–2, 1–3, and 1–4 cases) by the "unwanted terms" computed using the 1–2, 1–3 and 1–4 exclusion lists. This only requires a marginal addition of computing time because 1–2, 1–3, and 1–4 exclusion lists are typically 3% of the size of the neighbor list. This is easy to understand because one atom may have ~1000 neighbors within ~12 Å, but only forms bonds, angles, and dihedral angles with ~30 closest atoms.

MM atoms can use charges, induced dipoles, effective Gaussian potentials to interact with the electrons of QM atoms. These interactions are included as one-electron integrals of the basis functions when the QM electronic wavefunction is calculated. We have adapted various one-electron energy and energy derivative integral subroutines from GAMESS to the QuanPol program, and modified or redesigned parallel execution for them. Therefore, almost all MM and QM/MM calculations in QuanPol are parallelized. The distributed data interface (DDI)[115] and generalized DDI[116] are used for parallelization of QuanPol.

Neighbor list

Verlet neighbor list can be used to increase the execution efficiency of MD simulations.[83] QuanPol uses a simple cell-list scheme to generate a large neighbor list, which is typically two times larger than the small neighbor list and has a longer updating cycle (e.g., every 55 fs). The small list can be efficiently and frequently (e.g., every 11 fs) generated from the large list. QuanPol uses an automatic method to determine when to update a neighbor list. The atoms displace more than 0.2 and less than 0.9 of the buffer width are stored in 7 lists called fast-lists. When there are ~100 atoms in the fourth fast-list, which stores atoms that have displaces between 0.5 and 0.6 of the buffer width, it is fairly quick to check the pair distances between the atoms in all fast-lists. New atom pairs are added to the current list to avoid an immediate update, unless the number of atoms in the fourth fast-list exceeds a value like 300. For an equilibrium system, the frequencies of updating the large and small neighbor lists are almost constants. QuanPol identifies the frequencies and skips unnecessary checking of the fast-lists. For example, when buffer widths of 4.0 and 1.0 Å are used to generate the large and small neighbor lists, respectively, the lists update every ~55 and ~11 MD steps (time step size 1 fs) for a PBC system with 9121 protein atoms, 45 ions, and 60759 water atoms at $T = 310$ K, $P = 1$ bar, and cubic box size of ~88.8 Å. In this case, it is safe to skip the first 48 steps (estimated as 55 less the square root of 55) for the large list and the first eight steps for the small list. The fast-list method is inspired by and similar to the heuristic method[9,117,118] implemented in the CHARMM[7–9] program. Computer memory is usually not a problem for storing both the large and small neighbor lists. However, if memory is indeed a problem, the large and small lists can be merged and only one list will be used.

Bond length constraint through RATTLE

The RATTLE[89] method is implemented to constrain pairwise distances between atoms, typically atoms forming covalent bonds. RATTLE works together with both Beeman and velocity Verlet integration algorithms. We also implemented two geometry optimization algorithms that can use RATTLE to constrain distances (see subsection Geometry optimization). A RATTLE list is built according to user specifications in the input file. For example, a user may request to constrain all bond lengths found in the topology file, or only those bonds that contain at least one hydrogen atom, or any two atoms. For five-point water models, such as TIP5P and TIP5P-E,[119,120] to use the RATTLE method, real masses must be used for all points. In QuanPol, we name TIP5P and TIP5P-E atoms as O1, H2, H3, L4, and L5, with real masses 14.00, 1.08, 1.08, 1.00, and 1.00 atomic mass unit (amu), respectively. To enforce rigidity, nine independent pairs of distances are constrained for each TIP5P or TIP5P-E water molecule. Similarly, RATTLE in QuanPol can be applied to other rigid solvent molecules.

Force field

User defined force field can be specified from the input deck, and/or by supplying potential in a QuanPol data file for water and simple ions. Currently, nine water models are implemented in the QuanPol data file: TIP5P and TIP5P-E,[109,110]

Table 1. Some properties of liquid water at 298 K and 1 bar obtained from QuanPol MD simulation and different water models.

Model	Flexibility	Polarization	Density (g/cm³)	ΔH_{vap} (kcal/mol)	Dielectric constant	Self-diffusion constant (Å²/ns)
Experiment			0.99705	10.51	78	230
CHARMM TIP3P	Rigid	No	1.01522	10.42	105	546
TIP3P	Rigid	No	0.98584	10.15	102	559
TIP5P-E	Rigid	No	1.00547	10.39	92	290
SPC	Rigid	No	0.97830	10.54	75	435
SPC/E	Rigid	No	1.00139	11.75[a]	70	290
SPC/Fw	Flexible	No	1.00488	11.79	68	217
QP301	Flexible	No	0.99278	11.96	92	232
QP302	Flexible	Yes	0.99053	10.71	52	230

[a] After −1.25 kcal/mol "polarization correction" it is 10.50 kcal/mol.

SPC,[121] SPC/E,[122] SPC/Fw,[123] TIP3P[124] (both the original version and the CHARMM version), as well as the QP301 and QP302 models parameterized by us.

QuanPol can directly read some parameter and topology files compiled and distributed by the CHARMM, AMBER, OPLS-AA, and MMFF94 developers. MMFF94 is implemented in QuanPol for general organic molecules and some metal ions as described in the original MMFF94 papers, and tested with a full pass using the MMFF94 validation suit of 761 cases.[46–49] CHARMM is made available in QuanPol for amino acids, nucleic acids, and simple ions and tested using the relavent parameter and topology files in the packages named "toppar_c35b2_c36a2.tgz" and "toppar_c36_aug12.tgz" available at "http://mackerell.umaryland.edu/CHARMM_ff_params.html". AMBER is made available in QuanPol for amino acids, nucleic acids, and simple ions. OPLS-AA[44,45] parameters for amino acids and simple ions are made available by using the topology and parameter files from the CHARMM package "toppar_c35b2_c36a2.tgz". These force fields are not made available for general molecules. For example, the CHARMM new generalized force field CgenFF[125] for drug like molecules have not been made available. However, QuanPol is able to read the general AMBER force field (GAFF[125]) files in the mol2 format generated by Amber-Tools12[127], and read the GAFF parameter file "gaff.dat"[126] to establish the GAFF force field. Because the parameter and topology files compiled and distributed by developers of these standard force fields have consistent styles and formats, updated versions of these files can be directly used to perform QuanPol calculation with little or no modifications of the Quan-Pol source code. For example, QuanPol can use the AMBER topology and parameter files published from 1991 to 2012.

Numerical Results and Discussion

Water models

We used QuanPol to perform MD simulations of liquid water using five well-known three-point and five-point water models: SPC,[121] SPC/E,[122] SPC/Fw,[123] TIP3P[124] (both the original version and the CHARMM version), and TIP5P-E.[119,120] 512 water molecules were confined in a cubic PBC box with side lengths around 25 Å. Ewald summation was used to include

long-range charge–charge interactions. LJ interaction was switched in the range 10.0–12.0 Å using the atom-atom switching function implemented in QuanPol. The Berendsen thermostat and barostat were used to control the temperature and pressure with a time constant of 200 fs for temperature scaling and a time constant of 200 fs for pressure scaling, and with bath $T = 298.15$ K and bath $P = 1.0$ bar. For rigid water models, a time step size of 1 fs was used. For flexible water models, a time step size of 0.5 fs was used. In each case, the system was equilibrated for 1.0 ns, then followed by 1.0 ns production run. All simulations were run using the Beeman integration algorithm.

The key properties obtained from these simulations are listed in Table 1. Clearly, most of these results match well with the results in the literature. For example, the QuanPol TIP5P-E density 1.0055 g/cm³, enthalpy of vaporization 10.39 kcal/mol, dielectric constant 92, diffusion constant 290 Å²/ns are very close to, respectively, 1.0000 g/cm³, 10.38 kcal/mol, 92, and 280 Å²/ns reported by Rick.[120] The radial distribution functions were also obtained for these water models (data not shown). For the TIP5P-E model, the first peak is at 2.73 Å with a value of 2.86, in excellent agreement with the literature results.[120,128]

The enthalpy of vaporization (ΔH_{vap}) for the flexible SPC/Fw model is estimated to be 11.79 kcal/mol from the QuanPol simulation, significantly larger than the experimental value 10.51 and the value 10.72 reported by the authors of the SPC/Fw model.[123] It seems that the authors of the SPC/Fw model used the average potential energy in the MD and converted it to ΔH_{vap}.[123] If we do the same conversion here, the value is 10.90 kcal/mol, close to their value 10.72 kcal/mol. However, this assumes that the reference vapor state of a water molecule has zero internal potential energy. For a flexible three-point water model at 298.15 K, each of the three internal degrees of freedom has roughly 0.3 kcal/mol potential energy due to near-harmonic vibration. Considering the internal potential energy of vapor state water molecules, here the ΔH_{vap} is estimated to be 11.79 kcal/mol for the SPC/Fw model.

We also tested two three-point flexible water potentials using QuanPol. One is nonpolarizable, named QP301; one is polarizable with a dipole polarizability point at the oxygen site, named QP302. Both of them use equilibrium O—H bond

Table 2. Interaction potential parameters of the QP301 and QP302 three-point water models.

Model	QP301		QP302	
site	O	H	O	H
Lennard–Jones $R_{min}/2$ (Å)	1.7820	NA	1.8142	NA
Lennard–Jones ε (kcal/mol)	0.1520	NA	0.1520	NA
Dipole polarizability (Å^3)	NA	NA	0.8000	NA
Charge (e)	−0.8000	+0.4000	−0.7160	+0.3580
Effective Gaussian $a_1 \cdot exp(-b_1 R^2)$, a_1 (e/bohr)	30.0000	1.5000	30.0000	1.5000
Effective Gaussian $a_1 \cdot exp(-b_1 R^2)$, b_1 (bohr^{-2})	1.1200	2.0000	1.1500	2.0000
Effective Gaussian $a_2 \cdot exp(-b_2 R^2)$, a_2 (e/bohr)	−0.0330	−0.0125	−0.0440	−0.0144
Effective Gaussian $a_2 \cdot exp(-b_2 R^2)$, b_2 (bohr^{-2})	0.3000	0.3000	0.3000	0.3000
Two sets of effective Gaussian potentials are used for each site.				

length 1.03 Å and H—O—H bond angle 109.47 degree, and harmonic bond stretching and bending force constants 900 kcal/mol/Å^2 and 110 kcal/mol/rad^2, respectively. The intermolecular interaction potential parameters are listed in Table 2. The QP301 model uses −0.8000 and +0.4000 e charges for the oxygen and hydrogen sites. The QP302 uses a dipole polarizability of 0.8000 Å^3 for the oxygen site. The results in Table 1 show that the key properties of both QP301 and QP302 are comparable to other water models, except for that QP302 has a lower dielectric constant 52. It is very difficult to improve the dielectric constant when induced dipole polarization is used. In general, we found that when point charges and LJ potential are used, it is very difficult to calibrate an accurate induced dipole polarizable water model with either three sites or five sites. Because QP302 fulfills the minimum requirement of a flexible and polarizable water model, we used it in several demonstrative calculations in this article.

NaCl dissociation potential of mean force

We used two different methods implemented in QuanPol to compute the dissociation PMF of Na$^+$ and Cl$^-$ ion pair in water. The MD system contains a pair of Na$^+$ and Cl$^-$ ions and 854 TIP5P-E[119,120] water molecules in a cubic box with side length of ~29.3 Å. The CHARMM parameters for the Na$^+$ and Cl$^-$ ions were used. For Na$^+$ a positive unit charge and the LJ parameters $R_{min}/2 = 1.36375$ Å, $\varepsilon = 0.0469$ kcal/mol were used. For Cl$^-$ a negative unit charge and the LJ parameters $R_{min}/2 = 2.27$ Å, $\varepsilon = 0.15$ kcal/mol were used. We tested a few cases using Ewald summation and shifting function for long-range charge–charge interactions, and found essentially the same results for this system. Therefore, only the shifting function (QuanPol keyword ISHIFT = 4) was used because it is much more efficient. The MD was run with a time step size of 2 fs. The Berendsen thermostat and barostat were used with a temperature scaling time constant of 200 fs and a pressure scaling time constant of 200 fs, and with bath $T = 298.15$ K and bath $P = 1$ bar. All simulations were run using the Beeman integration algorithm.

The first method was single topology MD FEP. In this method, the MD was run with the Na$^+$ and Cl$^-$ ions fixed at their Cartesian coordinates, such as one at (−1.25 Å, 0, 0) and one at (+1.25 Å, 0, 0). Except for the fixation of these two ions, the MD was a normal NPT run. At every MD step, the Na$^+$ and Cl$^-$ ions were "perturbed" to new positions, such as (−1.30 Å, 0, 0) and (+1.30 Å, 0, 0). All other atoms and molecules (here only water) were not affected. Using this perturbed configuration, a single-point potential energy calculation was performed (actually the forces were also calculated but not used). The Na$^+$–Cl$^-$ interaction potential energy would decrease as their distance changes from 2.50 to 2.60 Å. However, the total interaction potential energy of the entire system could either decrease or increase, depending on how the two ions interact with water molecules, especially the nearby ones. The unperturbed and perturbed potential energies, E_R and E_{R+dR}, respectively, were used to calculate the Gibbs free energy difference between the two distances using Zwanzig[90] formula in the following form:

$$G_{R+dR} - G_R = -k_B T ln \left\langle exp\left(-\frac{E_{R+dR} - E_R}{k_B T} \right) \right\rangle \qquad (18)$$

A series of MD simulations were run using Na—Cl distance R fixed at 2.2, 2.3, …, 5.9 Å, with an increment of 0.1 Å. In each MD simulation, the Na—Cl distance perturbation $dR = 0.1$ Å was used. Each case was equilibrated for 0.1 ns, followed by a 1.0 ns production run. The relative free energy in each case was printed out in the output file. The resulted PMF curve (the value at 2.60 Å is arbitrarily set to be zero) is plotted in Figure 1.

The second method was umbrella sampling with a bias harmonic potential function added to the Na$^+$ and Cl$^-$ ion pair. The harmonic potential had an equilibrium distance R (e.g., 2.2 Å) and a force constant of 100 kcal/mol/Å^2. With this additional harmonic potential function, the MD was a normal NPT run. The Na—Cl distances stayed around R during the MD simulation. A series of MD simulations were performed for R from 2.2 to 5.9 Å with an increment of 0.1 Å. Each case was equilibrated for 0.1 ns, followed by a 1.0 ns production run. The QPWHA program was used to analyze the results and generate the PMF curve. The WHAM implemented in the QPWHA program only performs pure histogram analysis without the Jacobian correction.[129,130] Therefore, to generate the correct 1D PMF curve, the free energy must be corrected using the following formula:

$$G_R^{Jacob} - G_{R_0}^{Jacob} = k_B T ln\left(\frac{R^2}{R_0^2} \right) = 2k_B T ln\left(\frac{R}{R_0} \right) \qquad (19)$$

Here, R_0 is the reference Na—Cl distance at that the PMF value is set to be zero, that is, $R_0 = 2.60$ Å. At R_0 the volume free energy correction is also zero. Equation (19) can be easily obtained by considering the 1D "all-zero-PMF" for two ideal gas atoms. Defining the first atom as the origin of the coordinate, the second atom samples all the space with an even probability density. Therefore, in the umbrella sampling radial histogram, the counts are proportional to the square of R. A
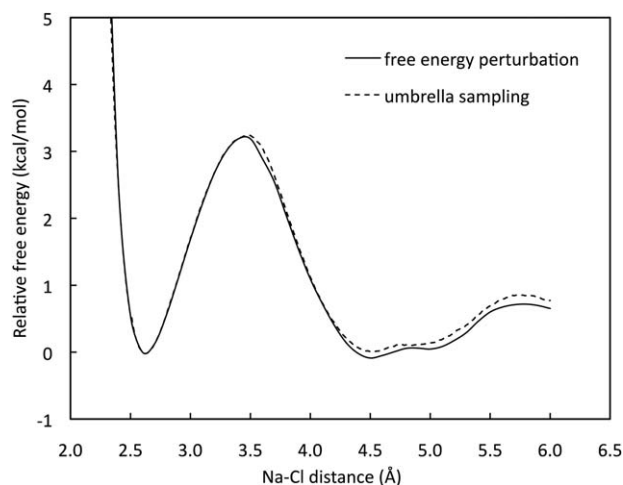
Figure 1. One-dimensional PMF (1D-PMF) obtained from QuanPol free energy simulation for Na^+Cl^- dissociation in aqueous solvent using pure MM method. The FEP method and umbrella sampling method produce virtually the same results.

direct application of the WHAM would simply produce a PMF curve that shows lower (more negative) free energy at larger distances. The Jacobian correction must be used to produce the 1D "all-zero-PMF" for two ideal gas atoms. In practice, only relative corrections can be applied by using eq. (19).

With the Jacobian correction, the 1D PMF from umbrella sampling is plotted in Figure 1 (the value at 2.60 Å is set to be zero). Clearly, this curve is very similar to the PMF curve from FEP calculations (Fig. 1). The differences might be caused by the incomplete sampling in both FEP and umbrella sampling, and might also be caused by the WHAM fitting algorithm. These results suggest that FEP and umbrella sampling can produce very similar results for this system. The Na^+Cl^- dissociation PMF curve shows the first minimum at 2.63 Å. The second and third minima are at 4.5 and 5.0 Å. The first and second maxima are at 3.46 and 4.9 Å. We note that the PMF curve is intrinsically determined by the force field parameters, which may not necessarily be of the highest accuracy.

QM/MMpol MD simulation

Energy-conserved MD simulation is the root of a MD program. To achieve energy conservation, the forces on all atoms must be evaluated accurately, and the MD integration algorithm must be a good one and implemented correctly. The analytic energy gradients for all QM/MMpol methods in QuanPol have been rigorously examined against numerical gradients. Usually the accuracy of the analytic gradients is better than 1.0×10^{-6} hartree/bohr.

We used QuanPol to perform Bohn–Oppenheimer style QM/MMpol MD simulations in constant volume and constant energy (NVE) ensembles for hydrogen cyanide (HCN) and 507 water molecules in a cubic PBC box with a side length of ~24.8 Å. The HCN molecule was described with RHF (ground state), B3LYP (ground state), TD-B3LYP (first singlet excited state S_1), RMP2 (ground state), ZAPT2 (first triplet excited state), GVB (first triplet exited state, T_1) and MCSCF [ground

state with a (4,4) active space] methods, and with the aug-cc-pVDZ basis set.[74] The water molecules were described with the three-point flexible and polarizable model QP302 (see Table 2). In these QM/MM simulations, no force field was used for the HCN molecule. The QM–MM LJ interaction typically used in QM/MM methods was replaced by the effective Gaussian potentials at the MM atoms. The water molecules carried charge, induced dipole, and effective Gaussian repulsion-dispersion potentials (see Table 2) that can directly interact with the electrons of the HCN molecule. The charges and induced dipoles of the water molecules also interact directly with the nuclei of the HCN molecule. The atom-atom switching function implemented in QuanPol was used for LJ, charge–dipole, and dipole–dipole interactions, with $r_a = 10$ Å and $r_b = 12$ Å; an atom-atom shifting function (QuanPol keyword ISHIFT = 1) was used for charge–charge interaction, with $r_b = 12$ Å. The QM-MM style switching function was used for QM–LJ, QM–charge, and QM–dipole interactions, with $r_a = 4$ Å and $r_b = 12$ Å. A time step size of 0.5 fs was used. In each case, 10,000 steps of QM/MM MD were run using the Beeman integration algorithm.

The total QM/MMpol energy (kinetic and potential, including QM energy) in the MD simulation is plotted in Figure 2. Clearly, the total energy is conserved very well in 10,000 steps (5 ps), with fluctuations in magnitudes of ~0.5 kcal/mol. Other tests show that when a time step size 0.1 fs is used, the total energy fluctuates only by ~0.02 kcal/mol.

Comparison between QuanPol and CHARMM results for proteins

We used the "CHARMM INterface and Graphics" website (www.charmming.org)[131] to perform several calculations for four small proteins to confirm that QuanPol can reproduce CHARMM potential energies. The reason that we used this website rather than directly using the CHARMM program is that we are not skilled users of CHARMM. The four small proteins are the Trp-cage miniprotein construct (PDB: 1L2Y),[132] the cysteine-rich terminal domain of Hydra minicollagen (PDB: 1SP7),[133] a copper-binding protein from fungus *Neurospora crassa* (PDB: 1T2Y),[134] and the long sarafotoxin srtx-i3 (PDB: 2LDE).[135] No water molecules were included in the calculation. The coordinates were downloaded from the CHARMMing website (which processed the original PDB files and added H atoms, and performed some steps of optimization for the H atoms) and used to prepare our QuanPol input files. Therefore, the atomic coordinates were the same (double precision accuracy). Each of these small proteins have 20–25 residues and ~300 atoms. Due to the requirement of manual editing of the coordinates, it is very difficult for us to test large proteins. Only the first models in these PDB files were used. Both QuanPol and CHARMM used the topology and parameter files "top_all27_prot_na.rtf" and "par_all27_prot_na.prm," including the CMAP term. In our QuanPol calculations, the CHARMM atom-atom shifting function (QuanPol keyword ISHIFT = 4) with cutoff = 12.0 Å for charge–charge interaction was used, and the atom-atom switching function implemented in
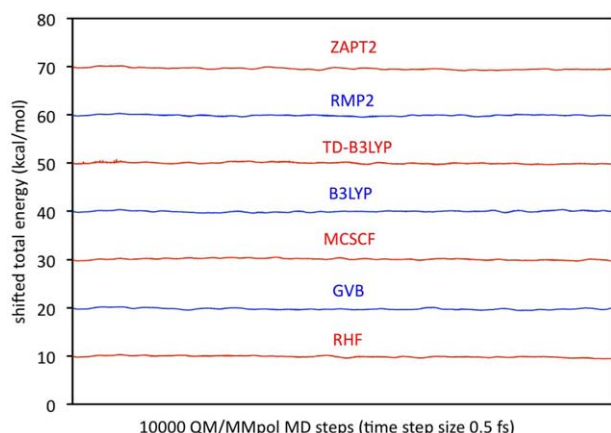
Figure 2. The total energies at every 10 MD steps in 10,000 steps of QM/MMpol style MD simulations for a hydrogen cyanide (HCN) molecule and 507 water molecules. Seven different QM methods with the aug-cc-pVDZ basis set are used for HCN and the three-point flexible and polarizable water model (QP302) is used for water. The energies are level-shifted by an arbitrary amount for each case so they appear in the same plot. In all the cases, the total energy fluctuates by ~0.5 kcal/mol, with a standard deviation of ~0.2 kcal/mol for all the 10,000 configurations.

QuanPol was used for LJ terms for the range 10.0–12.0 Å. The force field potential energies are listed in Table 3.

Clearly, QuanPol results match the CHARMM results very well. For bond, angle, dihedral rotation, dihedral bending terms, the results are identical. The differences in the CMAP terms are expected because in QuanPol we implemented the CMAP method independently using a different interpolation scheme. The small differences (all ~0.01 kcal/mol) in the LJ terms are also expected because CHARMM uses an atom-atom switching function that is different from the one implemented in QuanPol.

There are differences in the charge–charge interaction energies. Three cases are very small: +0.00517, +0.00107, +0.00445 kcal/mol. One case is large: +0.75194 kcal/mol. Because QuanPol and CHARMM used the same atom-atom shifting function, the QuanPol charge-charge energies should be identical to the CHARMM values. Indeed, in three cases, the energy differences are very small (+0.00517, +0.00107, and +0.00445 kcal/mol). The QuanPol values differ from CHARMM values by a scaling factor 0.9999762. Therefore, these small differences are caused by the use of slightly different physical constants and energy conversion factors in CHARMM and

QuanPol. Because QuanPol is implemented in GAMESS, we used the GAMESS values for these constants and conversion factors. The large difference +0.75194 kcal/mol in the 2LDE case is due to a different assignment of charges to the two amide H atoms of Asn13 in the 2LDE structure. CHARMMing website assigned a charge of 0.30 e to the cis H atom (cis to the amide O atom) of the Asn13 residue in 2LDE and a charge of 0.32 e to the $trans$ H atom ($trans$ to the amide O atom), and gave −482.33829 kcal/mol as the charge-charge energy for 2LDE. This assignment is not consistent with the CHARMM topology file "top_all27_prot_na.rtf," which states that the cis H atom of an Asn residue should have a charge of 0.32 e, and the $trans$ H atom should have a charge of 0.30 e. For the same 2LDE structure, CHARMMing website assigned a charge of 0.32 e to the cis H atom of Asn23 and a charge of 0.30 e to the $trans$ H atom, consistent with the topology file. Following the topology file "top_all27_prot_na.rtf," QuanPol assigned a charge of 0.32 e to the cis H atoms and a charge of 0.30 e to the $trans$ H atoms of both Asn13 and Asn23, and gave a charge-charge energy −481.58635 kcal/mol. We did a test by switching these two charges of Asn13 in the QuanPol input file, and obtained −482.32687 kcal/mol, a value essentially identical to the CHARMM value when corrected by the scaling factor 0.9999762.

We also tested various peptides, proteins, and DNA/RNA molecules for AMBER force field using QuanPol. Two independent sets of topology and parameters files were used: one set from the AMBER package, "all_amino94.in," "all_aminont94.in," "all_aminoct94.in," and "parm94.dat", and one set from the CHARMM package, "top_amber_cornell.inp" and "par_amber_cornell.inp." The QuanPol potential energies using these two sets of topology and parameter files always match to 10^{-7} kcal/mol for the tested proteins with hundreds to thousands atoms, and always match to 10^{-9} kcal/mol for the tested small DNA/RNA molecules.

IR spectrum of acetic acid

We used QM/MMpol style MP2 MD simulation method to compute IR spectrum of acetic acid in aqueous solution. The system consisted of one acetic acid molecule and 503 water molecules in a cubic PBC box with a side length of ~24.7 Å. The water molecules were modeled with the three-point

Table 3. Comparison between CHARMM force field energies (kcal/mol) computed using the QuanPol program and the CHARMM program for four small proteins.

	1L2Y (304 atoms)		1SP7 (352 atoms)		1T2Y (271 atoms)		2LDE (383 atoms)	
	CHARMM	QuanPol	CHARMM	QuanPol	CHARMM	QuanPol	CHARMM	QuanPol
Bond + Uray Bradley	40.06515	40.06515	36.37265	36.37265	69.09326	69.09326	32.40748	32.40748
Angle bending	73.74222	73.74222	61.05588	61.05588	43.46733	43.46733	31.71450	31.71450
Dihedral rotation	118.06877	118.06877	163.97550	163.97550	128.33697	128.33697	123.91744	123.91744
Dihedral bending	0.81990	0.81990	0.62454	0.62454	0.28170	0.28170	0.53894	0.53894
CMAP	−32.68845	−32.01436	−17.56636	−17.58863	−35.54992	−35.97501	−20.01132	−19.79101
LJ	−41.90223	−41.91244	51.86284	51.85082	396.97377	396.96451	22.73694	22.72537
Charge-charge	−217.95132	−217.94615	−44.80481	−44.80374	188.07425	188.06980	−482.33829[a]	−481.58635

[a] Due to a different charge assignment to the Asn13 residue in 2LDE. See text for discussion.

flexible and polarizable water model called QP302. The acetic acid molecule was modeled with the general AMBER force field (GAFF,[126] generated by using AmberTools12[127]) in pure MM calculation; it was modeled with the MP2/aug-cc-pVDZ[74] method in QM/MM calculation. In the MP2/QP302 simulations, no force field was used for the acetic acid molecule. The QM–MM LJ interaction typically used in QM/MM methods was replaced by the effective Gaussian potentials. The QP302 water molecules carried charge, induced dipole and effective Gaussian potentials (see Table 2) that can directly interact with the electrons of the acetic acid molecule. The charges and induced dipoles of the water molecules also interact directly with the nuclei of the acetic acid molecule. Using pure MM method, the system was equilibrated for 0.2 ns (200,000 steps with $\Delta t = 1$ fs) with the Berendsen thermostat and barostat (both the temperature and volume scaling time constants were 200 fs) and with bath $T = 298.15$ K and bath $P = 1$ bar. Then 10,000 steps of MP2/QP302 MD simulation were run using the last configuration of the MM equilibration run to equilibrate the system at the QM/MM level. Finally 15,000 steps of MP2/QP302 production MD were run. The atom-atom switching function implemented in QuanPol was used for LJ, charge–dipole, and dipole–dipole interactions, with $r_a = 10$ Å and $r_b = 12$ Å; an atom-atom shifting function (QuanPol keyword ISHIFT = 4) was used for charge–charge interaction, with $r_b = 12$ Å. The QM-MM style switching function was used for QM-LJ, QM-charge and QM-dipole interactions, with $r_a = 8$ Å and $r_b = 12$ Å. The dipole moment of the QM atoms, the dipole moment of all MM atoms, and the total dipole moment of the entire QM/MM system, were printed out at every MD step. All simulations were run using the Beeman integration algorithm.

The parallel MP2 program implemented by Ishimura et al.[107,108] was used together with the QuanPol routines that add induced dipoles to the MP2 method.[59] The MP2/QP302 MD simulation was run using an eight-node Linux cluster with a gigabit network. Each node has two Dual Core AMD Opteron 275 Processors (four cores per node and total 32 cores) running at 2.2 GHz. It took 5.8 days to finish the 15,000 steps of MD simulation.

The IR spectra of gas phase acetic acid (monomer) were calculated with two independent methods. One is a straightforward Hessian harmonic vibration analysis using the gas phase MP2/aug-cc-pVDZ[74] method. The other is a gas-phase energy-conserved MD simulation (15,000 steps with step time size of 1 fs) of one acetic acid molecule using the same MP2/aug-cc-pVDZ[74] method. The MD was performed using QuanPol with null MM atoms. For such a small system, temperature scaling is meaningless because the instantaneous temperature fluctuates very large, typically 100 K. This is the reason why we used energy conserved MD. It turned out that the average temperature in the MD simulation was 323 K. Because the IR spectrum is not sensitive to the temperature in the MD, we did not try to make the temperature closer to 298.15 K. The total energy (QM electronic energy with nuclear repulsion and nuclear kinetic) was conserved within ± 0.1 kcal/mol over the entire trajectory. The parallel MP2 program implemented by

Ishimura et al.[107,108] was used together with the QuanPol routines. The MP2 MD simulation was run using a six-node Linux cluster with a gigabit network. Each node has an Intel Core2 Quad Q9550 CPU (four cores per node and total 24 cores) running at 2.83 GHz. It took 2.7 days to finish the 15,000 steps of MP2 MD simulation.

The QPVIB was used to read the dipole moments from the 15,000-step production runs to generate the IR spectrum. The time correlation function was formulated to be a 2048-point discrete function over the range of 2048 fs so the resolution of the Fourier transformed frequency spectra is 16.28 cm^{-1}. The computed solution-phase IR spectrum and the gas-phase IR line spectrum of acetic acid are plotted in Figure 3.

Marechal reported seven IR absorption maximum frequencies (988, 1075, 1178, 1280, 1395, 1788, and 3583 cm^{-1}) for acetic acid monomer in the gas phase.[136] The gas-phase MP2 Hessian harmonic model (with no scaling) frequencies (995, 1056, 1204, 1328, 1403, 1792, 3741 cm^{-1}) are in good agreement with experiment. It is well known that MP2 harmonic frequencies are a few percent higher than experimental values. The gas-phase MP2 MD-simulated frequencies (993, 1059, 1189, 1270, 1401, 1792, 3779 cm^{-1}) are very similar to the harmonic model results.

Max and Chapados obtained IR spectrum of acetic acid in aqueous solution.[137] They reported the following band positions: 886 (weak), 1015 (medium), 1050 (weak), 1271 (strong), 1388 (strong), and 1706 (strong) cm^{-1}. The solution phase MP2/QP302 MD simulation gives 896 (weak), 1042 (medium), 1091 (weak), 1303 (strong), 1384 (strong), 1759 (strong) cm^{-1}, all in agreement with experiment.

The IR spectrum of acetic acid changes from the gas phase to aqueous solution phase. The most significant change is the blue shift of the v_{C-O} mode from 1178 to 1271 cm^{-1}. This is due to the solute–solvent dynamic interaction, and can hardly be explained without using explicit solvent MD simulations. For example, we performed a Hessian harmonic calculation using MP2/aug-cc-pVDZ and the FixSol[68] continuum model, and found that the v_{C-O} mode shows a frequency of 1209 cm^{-1} in solution, almost identical to the gas phase harmonic value 1204 cm^{-1}. The MP2/QP302 MD simulation used here is able to accurately reproduce this blue shift (1189 cm^{-1} in gas and 1303 cm^{-1} in solution).

Conclusion

The quantum chemistry polarizable force field program (QuanPol), which seamlessly integrates a full spectrum of QM and MM methods to perform QM/MM and QM/MM/continuum style MD simulations, has been implemented and tested. As of August 2013, QuanPol has more than 65,600 lines of Fortran code. Here, we summarize the features of the QuanPol program:

1. The QM methods can be HF, DFT, GVB, MCSCF, MP2, and TDDFT. The MM force field methods can be user specified, or a standard force field such as MMFF94, CHARMM, AMBER, and OPLS-AA. Induced dipole polarizable force fields can be used. The FixSol-induced surface charge continuum solvation model can be used together with all
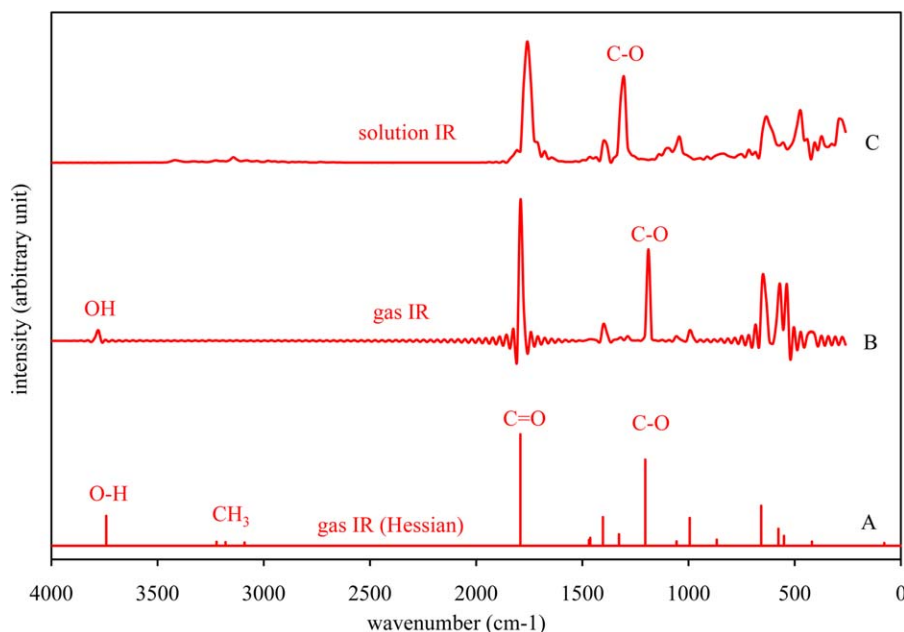
Figure 3. A) The gas phase IR line spectrum from an MP2 Hessian harmonic model calculation (the heights of all lines are increased for better visibility); B) The gas-phase IR spectrum from MP2 MD simulation; C) The IR spectrum of acetic acid in aqueous solution obtained from MP2/QP302 MD simulation.

the QM/MM combinations. The induced dipoles of the MM atoms and the induced surface charges of the continuum solvation model are self-consistently and variationally determined together with the QM wavefunction.

2. Analytic gradients for all of these methods are implemented so geometry optimization and MD simulation can be performed. Dynamic properties such as IR spectrum can be obtained by using QuanPol QM/MM MD simulation methods for ground state and excited state molecules in solution or proteins.

3. A capping H atom method is implemented to treat QM and MM covalent boundaries. This method is similar to the "link H atom" scheme, but does not introduce additional degrees of freedom. It is also similar to the "boundary atom" scheme, but does not need sophisticated effective or pseudopotentials.

4. MD FEP and umbrella sampling methods are implemented for MM and QM/MM methods. Properties such as free energy profile (PMF), binding free energy, acid ionization constant (pKa), and reduction potential can be obtained using QuanPol and its service programs. Both MM and QM/MM "alchemical" FEP schemes can be applied.

5. QuanPol can use a unique average position mean field QM/⟨MM⟩ MD simulation method to accelerate QM/MM calculations by 100–1000 times, bringing the QM/MM computational costs down to a level similar to pure MM methods. MD FEP methods are implemented for this mean field QM/⟨MM⟩ method. Properties such as free energy profile (PMF), binding free energy, acid ionization constant (pKa), and reduction potential can be efficiently computed.

QuanPol is integrated in GAMESS. A test version was released in August 2011 together with GAMESS. A new version of QuanPol with all of the functionalities described in this arti-

cle will be included into GAMESS and distributed free of charge to licensed GAMESS users. It does not require a separate license to use QuanPol. It is our hope that QuanPol can serve as a useful platform for comparison, application, and further development of different QM/MM methods.

Acknowledgments

[1] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227.

[2] T. E. Cheatham, M. A. Young, *Biopolymers* **2000**, *56*, 232.

[3] P. Cieplak, J. Caldwell, P. Kollman, *J. Comput. Chem.* **2001**, *22*, 1048.

[4] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.

[5] Z. X. Wang, W. Zhang, C. Wu, H. X. Lei, P. Cieplak, Y. Duan, *J. Comput. Chem.* **2006**, *27*, 781.

[6] J. W. Ponder, D. A. Case, D. Valerie, In Advances in Protein Chemistry, Vol. *66*; Academic Press, **2003**; 27–85.

[7] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187.

[8] A. D. MacKerell, Jr., B. Brooks, C. L. Brooks, III, L. Nilsson, B. Roux, Y. Won, M. Karplus, In The Encyclopedia of Computational Chemistry, Vol. *1*; P. v. R. Schleyer, Ed.;Wiley: Chichester, **1998**.

[9] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.

[10] W. L. Jorgensen, J. Tirado–Rives, *J. Comput. Chem.* **2005**, *26*, 1689.

[11] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **1994**, *91*, 43.

[12] E. Lindahl, B. Hess, D. van der Spoel, *J. Mol. Model.* **2001**, *7*, 306.

[13] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.

[14] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435.

[15] J. W. Ponder, F. M. Richards, *J. Comput. Chem.* **1987**, *8*, 1016.

[16] J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox. Gaussian 09, Revision a.1, Gaussian, Wallingford CT, **2009**.

[17] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347.

[18] M. S. Gordon, M. W. Schmidt, In Theory and Applications of Computational Chemistry; C. E. Dykstra, G. Frenking, K. S. Kim, G. E. Scuseria, Eds.; Elsevier: Amsterdam, **2005**.

[19] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Yeh Lin, T. Van Voorhis, S. Hung Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. -P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. Min Rhee, J. Ritchie, E. Rosta, C. David Sherrill, A. C. Simmonett, J. E. Subotnik, H. Lee Woodcock Iii, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer Iii, J. Kong, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.

[20] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931.

[21] M. F. Guest, I. J. Bush, H. J. J. Van Dam, P. Sherwood, J. M. H. Thomas, J. H. Van Lenthe, R. W. A. Havenith, J. Kendrick, *Mol. Phys.* **2005**, *103*, 719.

[22] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong, *Comput. Phys. Commun.* **2010**, *181*, 1477.

[23] P. E. Blochl, *Phys. Rev. B* **1994**, *50*, 17953.

[24] M. E. Casida, In Recent Advances in Density Functional Methods; D. P. Chong, Eds.; World Scientific: Singapore, **1995**, pp. 155.

[25] M. E. Casida, C. Jamorski, K. C. Casida, D. R. Salahub, *J. Chem. Phys.* **1998**, *108*, 4439.

[26] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.

[27] F. J. Vesely, *J. Comput. Phys.* **1977**, *24*, 361.

[28] M. Neumann, F. J. Vesely, O. Steinhauser, P. Schuster, *Mol. Phys.* **1978**, *35*, 841.

[29] F. H. Stillinger, C. W. David, *J. Chem. Phys.* **1978**, *69*, 1473.

[30] P. Barnes, J. L. Finney, J. D. Nicholas, J. E. Quinn, *Nature* **1979**, *282*, 459.

[31] A. Warshel, *J. Phys. Chem.* **1979**, *83*, 1640.

[32] V. Luzhkov, A. Warshel, *J. Am Chem. Soc.* **1991**, *113*, 4491.

[33] A. Warshel, Z. T. Chu, *J. Phys. Chem. B* **2001**, *105*, 9857.

[34] J. Gao, *J. Am. Chem. Soc.* **1994**, *116*, 9324.

[35] J. Gao, K. Byun, *Theor. Chem. Acc.* **1997**, *96*, 151.

[36] J. L. Gao, *J. Comput. Chem.* **1997**, *18*, 1061.

[37] M. A. Thompson, G. K. Schenter, *J. Phys. Chem.* **1995**, *99*, 6374.

[38] M. A. Thompson, *J. Phys. Chem.* **1996**, *100*, 14492.

[39] L. Jensen, P. T. van Duijnen, J. G. Snijders, *J. Chem. Phys.* **2003**, *119*, 3800.

[40] C. B. Nielsen, O. Christiansen, K. V. Mikkelsen, J. Kongsted, *J. Chem. Phys.* **2007**, *126*, 18.

[41] T. D. Poulsen, J. Kongsted, A. Osted, P. R. Ogilby, K. V. Mikkelsen, *J. Chem. Phys.* **2001**, *115*, 2393.

[42] Y. Lin, J. Gao, *J. Chem. Theory Comput.* **2007**, *3*, 1484.

[43] S. Yoo, F. Zahariev, S. Sok, M. S. Gordon, *J. Chem. Phys.* **2008**, *129*, 8.

[44] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.

[45] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474.

[46] T. A. Halgren, R. B. Nachbar, *J. Comput. Chem.* **1996**, *17*, 587.

[47] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 520.

[48] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 553.

[49] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 616.

[50] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

[51] W. A. Goddard, T. H. Dunning, W. J. Hunt, P. J. Hay, *Acc. Chem. Res.* **1973**, *6*, 368.

[52] P. E. M. Siegbahn, J. Almlof, A. Heiberg, B. O. Roos, *J. Chem. Phys.* **1981**, *74*, 2384.

[53] J. Ivanic, K. Ruedenberg, *J. Comput. Chem.* **2003**, *24*, 1250.

[54] J. Ivanic, K. Ruedenberg, *Theor. Chem. Acc.* **2001**, *106*, 339.

[55] J. M. Rintelman, M. S. Gordon, G. D. Fletcher, J. Ivanic, *J. Chem. Phys.* **2006**, *124*, 5.

[56] H. Li, *J. Chem. Phys.* **2009**, *131*, 184103.

[57] Y. Wang, H. Li, *J. Chem. Phys.* **2010**, *133*, 034108.

[58] D. J. Si, H. Li, *J. Chem. Phys.* **2011**, 135, 144107.

[59] H. Li, *J. Phys. Chem. A* **2011**, *115*, 11824.

[60] D. Si, H. Li, *J. Chem. Phys.* **2010**, *133*, 144,112.

[61] A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400.

[62] R. E. Tuzun, D. W. Noid, B. G. Sumpter, *Macromol. Theory Simul.* **1996**, *5*, 771.

[63] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, D. Cohen, *J. Chem. Phys.* **1996**, *105*, 1968.

[64] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodoscek B. R. Brooks, *J. Chem. Phys.* **2002**, *117*, 10,534.

[65] A. Klamt, G. Schuurmann, *J. Chem. Soc.Perkin Trans* **1993**, *2*, 799.

[66] H. Li, J. H. Jensen, *J. Comput. Chem.* **2004**, *25*, 1449.

[67] V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995.

[68] N. M. Thellamurege, H. Li, *J. Chem. Phys.* **2012**, *137*, 246,101.

[69] S. Chalmet, D. Rinaldi, M. F. Ruiz-López, *Int. J. Quantum Chem.* **2001**, *84*, 559.

[70] Q. Cui, *J. Chem. Phys.* **2002**, *117*, 4720.

[71] P. Bandyopadhyay, M. S. Gordon, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **2002**, *116*, 5023.

[72] H. Li, C. S. Pomelli, J. H. Jensen, *Theor. Chem. Acc.* **2003**, *109*, 71.

[73] R. H. Hertwig, W. Koch, *Chem. Phys. Lett.* **1997**, *268*, 345.

[74] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.

[75] Y. Zhang, T. -S. Lee, W. Yang, *J. Chem. Phys.* **1999**, 110, 46.

[76] Y. Zhang, *J. Chem. Phys.* **2005**, *122*, 024114.

[77] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.

[78] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer and C. Lennartz, *J. Mol. Struct.* **2003**, *632*, 1.

[79] H. Lin, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 3991.

[80] P. P. Ewald, *Ann. Phys.* **1921**, *369*, 253.

[81] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.

[82] Y. Zhou, M. Cook, M. Karplus, *Biophys. J.* **2000**, *79*, 2902.

[83] L. Verlet, *Phys. Rev.* **1967**, *159*, 98.

[84] D. Beeman, *J. Comput. Phys.* **1976**, *20*, 130.

[85] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *The J. Chem. Phys.* **1984**, *81*, 3684.

[86] H. C. Andersen, *J. Chem. Phys.* **1980**, *72*, 2384.

[87] H. C. Andersen, *J. Comput. Phys.* **1983**, *52*, 24.

[88] M. Frigo, S. G. Johnson, *Proc. IEEE* **2005**, *93*, 216.

[89] F. Jensen, *J. Chem. Phys.* **1995**, *102*, 6706.

[90] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.

[91] D. A. Pearlman, *J. Phys. Chem.* **1994**, *98*, 1487.

[92] S. Boresch, M. Karplus, *J. Phys. Chem. A* **1998**, *103*, 119.

[93] W. Yang, R. Bitetti-Putzer, M. Karplus, *J. Chem. Phys.* **2004**, *120*, 9450.

[94] M. R. Reddy, U. C. Singh, M. D. Erion, *J. Am. Chem. Soc.* **2004**, *126*, 6224.

[95] G. M. Torrie, J. P. Valleau, *Chem. Phys. Lett.* **1974**, *28*, 578.

[96] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187.

[97] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 1011.

[98] M. L. Sanchez, M. A. Aguilar, F. J. Olivares del Valle, *J. Comput. Chem.* **1997**, *18*, 313.

[99] M. L. S. Mendoza, M. A. Aguilar, F. J. Olivares del Valle, *J. Mol. Struct. THEOCHEM* **1998**, *426*, 181.

[100] I. F. Galván, M. L. Sánchez, M. E. Martín, F. J. Olivares del Valle, M. A. Aguilar, *Comput. Phys. Commun.* **2003**, *155*, 244.

[101] E. Rosta, M. Haranczyk, Z. T. Chu, A. Warshel, *J. Phys. Chem. B* **2008**, *112*, 5680.

[102] H. Nakano, T. Yamamoto, *J. Chem. Phys.* **2012**, *136*, 134,107.

[103] F. Cui, H. Li, *J. Chem. Phys.* **2013**, *138*, 174,114.

[104] H. Hu, Z. Lu, W. Yang, *J. Chem. Theory Comput.* **2007**, *3*, 390.

[105] M. J. Frisch, M. Head-Gordon, J. A. Pople, *Chem. Phys. Lett.* **1990**, *166*, 275.

[106] G. D. Fletcher, M. W. Schmidt, M. S. Gordon, *Adv. Chem. Phys.* **1999**, *110*, 267.

[107] K. Ishimura, P. Pulay, S. Nagase, *J. Comput. Chem.* **2006**, *27*, 407.

[108] K. Ishimura, P. Pulay, S. Nagase, *J. Comput. Chem.* **2007**, *28*, 2034.

[109] G. D. Fletcher, M. S. Gordon, R. S. Bell, *Theor. Chem. Acc.* **2002**, *107*, 57.

[110] C. M. Aikens, G. D. Fletcher, M. W. Schmidt, M. S. Gordon, *J. Chem. Phys.* **2006**, *124*, 014107.

[111] C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, M. S. Gordon, *Theor. Chem. Acc.* **2003**, *110*, 233.

[112] C. M. Aikens, M. S. Gordon, *J. Phys. Chem. A* **2004**, *108*, 3103.

[113] M. Chiba, T. Tsuneda, K. Hirao, *J. Chem. Phys.* **2006**, *124*, 144106.

[114] M. Chiba, T. Tsuneda, K. Hirao, *Chem. Phys. Lett.* **2006**, *420*, 391.

[115] G. D. Fletcher, M. W. Schmidt, B. M. Bode, M. S. Gordon, *Comput. Phys. Commun.* **2000**, *128*, 190.

[116] D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, S. Koseki, *J. Comput. Chem.* **2004**, *25*, 872.

[117] S. M. Thompson, *CCP5 Newsl.* **1983**, *8*, 20.

[118] D. Fincham, B. J. Ralston, *Comput. Phys. Commun.* **1981**, *23*, 127.

[119] M. W. Mahoney, W. L. Jorgensen, *J. Chem. Phys.* **2000**, *112*, 8910.

[120] S. W. Rick, *J. Chem. Phys.* **2004**, *120*, 6085.

[121] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunstere, J. Hermans, Intermolecular Forces; Reidel: Dordrecht, **1981**.

[122] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **1987**, *91*, 6269.

[123] Y. Wu, H. L. Tepper, G. A. Voth, *J. Chem. Phys.* **2006**, *124*, 024503.

[124] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.

[125] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, *J. Comput. Chem.* **2010**, *31*, 671.

[126] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.

[127] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, University of California, San Francisco, **2012**.

[128] T. Head-Gordon, G. Hura, *Chem. Rev.* **2002**, *102*, 2651.

[129] I. V. Khavrutskii, J. Dzubiella, J. A. McCammon, *J. Chem. Phys.* **2008**, *128*, 044106.

[130] J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932.

[131] B. T. Miller, R. P. Singh, J. B. Klauda, M. Hodoscek, B. R. Brooks, H. L. Woodcock, *J. Chem. Inform. Model.* **2008**, *48*, 1920.

[132] J. W. Neidigh, R. M. Fesinmeyer, N. H. Andersen, *Nat. Struct. Biol.* **2002**, *9*, 425.

[133] S. Meier, D. Häussinger, E. Pokidysheva, H. P. Bächinger, S. Grzesiek, *FEBS Lett.* **2004**, *569*, 112.

[134] P. A. Cobine, R. T. McKay, K. Zangger, C. T. Dameron, I. M. Armitage, *Eur. J. Biochem.* **2004**, *271*, 4213.

[135] G. Mourier, M. Hajj, F. Cordier, A. Zorba, X. Gao, T. Coskun, A. Herbet, E. Marcon, F. Beau, M. Delepierre, F. Ducancel, D. Servent, *Biochimie* **2012**, *94*, 461.

[136] Y. Marechal, *J. Chem. Phys.* 87, 6344.

[137] J. -J. Max, C. Chapados, *J. Phys. Chem. A* **2004**, *108*, 3324.