

Article

X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water

Wangshen Xie, Modesto Orozco, Donald G. Truhlar, and Jiali Gao

J. Chem. Theory Comput., **2009**, 5 (3), 459-467 • DOI: 10.1021/ct800239q • Publication Date (Web): 17 February 2009

Downloaded from <http://pubs.acs.org> on March 13, 2009

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)



ACS Publications
High quality. High impact.

X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water

Wangshen Xie,[†] Modesto Orozco,^{*,‡} Donald G. Truhlar,^{*,†} and Jiali Gao^{*,†,‡}

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis Minnesota 55455-0431, and Barcelona SuperComputer Center—Institute for Research in Biomedicina Joint Program on Computational Biology, C/ Jordi Girona 29, 08034 Barcelona, Spain and Josep Samitier 1-5, 08028 Barcelona, Spain

Received June 21, 2008

Abstract: A recently proposed electronic structure-based force field called the explicit polarization (X-Pol) potential is used to study many-body electronic polarization effects in a protein, in particular by carrying out a molecular dynamics (MD) simulation of bovine pancreatic trypsin inhibitor (BPTI) in water with periodic boundary conditions. The primary unit cell is cubic with dimensions $\sim 54 \times 54 \times 54 \text{ \AA}^3$, and the total number of atoms in this cell is 14281. An approximate electronic wave function, consisting of 29026 basis functions for the entire system, is variationally optimized to give the minimum Born–Oppenheimer energy at every MD step; this allows the efficient evaluation of the required analytic forces for the dynamics. Intramolecular and intermolecular polarization and intramolecular charge transfer effects are examined and are found to be significant; for example, 17 out of 58 backbone carbonyls differ from neutrality on average by more than 0.1 electron, and the average charge on the six alanines varies from -0.05 to $+0.09$. The instantaneous excess charges vary even more widely; the backbone carbonyls have standard deviations in their fluctuating net charges from 0.03 to 0.05, and more than half of the residues have excess charges whose standard deviation exceeds 0.05. We conclude that the new-generation X-Pol force field permits the inclusion of time-dependent quantum mechanical polarization and charge transfer effects in much larger systems than was previously possible.

1. Introduction

Molecular dynamics simulation has become a powerful tool for studying biochemical properties ranging from protein and nucleic acid dynamics and structural prediction to chemical reactions in enzymes.¹ At the heart of these calculations is the potential energy function that describes intermolecular interactions in the system,^{2–8} and often it is the accuracy of the potential energy surface (or its gradient field, called the force field) that determines the reliability of simulation results. Because of the size of condensed-phase systems and the complexity of biomacromolecules, one typically uses

molecular mechanics force fields, in which the potential energy surface for a macromolecular system is approximated by analytical functions describing bond stretches, bond angle bends, torsions, and nonbonded van der Waals and Coulomb interactions.^{2,3} The computational efficiency of analytic molecular mechanics force fields allows molecular dynamics simulations of biopolymer systems to be carried out with the extensive sampling required for rare event simulation, and classical simulations may be extended for long time scales and to large molecular systems. There is therefore considerable effort being expended to improve the physical representation and accuracy of such force fields, with special emphasis on including polarization effects to better represent electrostatic forces.^{3,4}

The development of molecular mechanics dates back to early studies of steric effects of organic compounds,^{5–7} and

* Corresponding author e-mail: modesto@mmb.pcb.ub.es (M.O.), truhlar@umn.edu (D.G.T.), gao@jialigao.org (J.G.).

[†] University of Minnesota.

[‡] Barcelona SuperComputer Center.

the basis of the current generation of force fields for biomolecular systems was established in the 1960s.^{2,8} The first molecular dynamics simulation of a protein was reported in 1977 by McCammon, Gelin, and Karplus;⁹ it lasted 8.8 ps for a small protein, in particular, bovine pancreatic trypsin inhibitor (BPTI) in the gas phase. The authors noted two limitations in that study. The first “is the approximate nature of the potential energy function,” which was essentially of the same type as we are using today, and the second “is the neglect of solvent”. Although a relatively short simulation was performed on a dry protein, the 1977 article⁹ is one of the classic studies in our field because of its vision that paved the way for molecular simulation and modeling as we know it today. Of course, solvation effects are now recognized as an unavoidable essential element in dynamics simulations because they play an inescapable role in biomolecular function, and tremendous progress has been made in the accuracy of conventional force fields for modeling biopolymers.^{3,10–15} Yet, it is sobering to notice that the physical representation and functional forms used in molecular mechanics have hardly changed.

Despite the success of molecular mechanics in biomacromolecular modeling, there are also shortcomings, such as inapplicability to chemical reactions and lack of polarization. Recognition of the latter has motivated many efforts^{3,4,16–33} to parametrize nonadditive polarization effects in the force fields. An alternative is to make a fundamental paradigm change in the functional form of the force fields and the representation of biomolecular systems,^{31–33} moving beyond the present classical development. With this motivation, we have introduced an explicit polarization (X-Pol) model based on quantum mechanics as a framework for a next-generation force field.^{34–36} In what follows, we report a molecular dynamics simulation of a fully solvated BPTI protein that, although short, employs an explicit quantal force field. This force field is currently represented by an available semiempirical method, namely the AM1 approximate molecular orbital theory. Nevertheless, the present study demonstrates the feasibility of an entirely new concept in force field development for large-scale simulation.

Polarization and charge transfer are intrinsic properties of the electronic structure of a molecular system, resulting in polar bonds, nonunit charge on functional groups, and electric response of electron density to an external field. Although molecular polarization is a well-defined property, its incorporation into an MM force field is not unique.^{3,4,16–30} Consequently, numerous models have been proposed for the classical treatment of polarization effects, and their validity is a subject of ongoing validation. In the X-Pol potential,^{31–36} the internal energy terms and electrostatic potentials used in the force field are described explicitly by a quantum chemical wave function. Since molecular polarization and charge transfer are represented naturally by electronic structure theory, no polarization terms need to be added, and hence there is no ambiguity in the choice of functional form for polarization terms or in the selection of internal degrees of freedom to define these terms. Furthermore, such a method can be used to model chemical reactions.

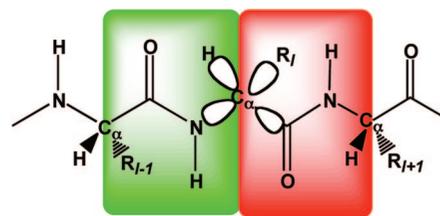


Figure 1. Definition of peptide units and the division of a protein into fragments at the C_{α} boundary atom. Two quantum mechanical fragments are highlighted in green and red, respectively, corresponding to residues $l - 1$ and l .

In the X-Pol potential, a molecular system is partitioned into fragments, such as an individual solvent molecule or a peptide unit or a group of such entities. The electronic interaction within each fragment is treated using electronic structure theory, while the interfragment electrostatic interactions are treated^{31,34–36} using a quantal analog of the combined quantum mechanical and molecular mechanical (QM/MM) approach (hence these interaction terms are sometimes called electrostatic QM/MM terms). Because the wave function of the entire system (which is assumed to be a Hartree product of antisymmetrized fragment wave functions) is variationally optimized,³⁵ we can take advantage of analytic gradient theory³⁷ to develop efficient methods for evaluating the contributions of internal energies and electrostatic interactions to forces.³⁵ Exchange repulsion and dispersion-like attraction between fragments are added by pairwise additive functions.

The X-Pol potential has been tested and applied to the simulation of liquid water³² and liquid hydrogen fluoride,³³ and it has been recently extended to treat fragments that are covalently bonded to one another.^{34–36} Here we employ analytic gradients to carry out the dynamical simulation of a small protein, in particular BPTI, in a water box with a size of about $54 \times 54 \times 54 \text{ \AA}^3$; the water box contains 4461 water molecules and one copy of the protein. We will analyze the polarization of the charge distribution of the protein and its significance for the description of protein–solvent interactions. The stability of MD simulations using the X-Pol potential for macromolecular simulations in water will also be demonstrated. The calculations presented here are designed to set the stage for systematically parametrizing the X-Pol potential to achieve the goal of chemical accuracy in such simulations.

In Section 2 we briefly review the theory of the X-Pol potential, and section 3 gives computational details. Sections 4–6 present results, timings, discussion, and comments on future prospects.

2. Theoretical Background

The design of the X-Pol potential has been described in a series of publications,^{31–36} and here we only present the necessary background for the simulation of a solvated protein. We adopt the peptide unit convention as defined by IUPAC,³⁸ although we note that the residue convention is typically used in other force fields.^{3,10–14} As endorsed in the IUPAC rules,³⁸ we will refer to a peptide unit by its residue name. Figure 1 shows the division of a peptide chain

into peptide units at a C_α carbon; each peptide unit is defined as a quantum mechanical (QM) fragment in the present calculation, and the C_α atom is called a boundary atom. Only the valence electrons of the boundary atom are treated explicitly, so the effective nuclear charge is four and the associated number of electrons is also four. Both the nuclear charge and electrons are divided equally into the two neighboring fragments. The boundary atom has four hybrid bonding orbitals,^{34,39} such that each fragment has two of them as active orbitals and the other two as auxiliary orbitals. This partition of a polypeptide results in two “pseudo atoms”, which have identical coordinates, and each of which is half of a boundary C_α atom.^{34–36}

The wave function of the entire system is written as a Hartree product of Slater determinant wave functions of individual fragments.^{31,34,35} In the X-Pol potential, the internal energies of the fragments are treated with electronic structure theory, and the interactions between fragments are described by a combined quantum mechanical and molecular mechanical (QM/MM^{40–42}) approach. No bond stretching, bending, or torsion terms appear because such interactions are represented by quantum mechanics, and no harmonic assumptions or analytic anharmonicity terms appear. It should be emphasized that at the top of the hierarchy of approximations in the X-Pol model, there is no restriction on the level of molecular orbital theory or density functional theory used to represent each individual fragment. In principle, it is possible to use a high-level quantum model to treat the region of interest and to use a lower level of theory to represent the rest of the system. In the present version of the method, the electronic structure calculations are carried out by valence-only semiempirical molecular orbital theory with the neglect of diatomic differential overlap (NDDO⁴³). Thus the electronic wave function includes only valence electrons, and core electrons are combined with nuclei and treated as frozen atomic cores.

The total energy of the system includes the electronic energies of the fragments (each including half of the electronic Coulombic interactions and half of the core interactions to avoid double counting) plus an empirical van der Waals term. Thus,

$$E_{tot} = E_{elec} + E_{vdw} \quad (1)$$

where the van der Waals energy term is required because the electronic structure calculation omits electron correlation and exchange repulsion between electrons in different fragments. The van der Waals term is a sum^{32,35} of Lennard-Jones potentials, including both repulsion due to exchange and dispersion-like attraction due to medium-range correlation energy. Note that Lennard-Jones interactions are omitted for atom pairs within the same fragment and for those separated by less than 3 bonds (*i.e.*, 1–2 and 1–3 van der Waals interactions are excluded) as in most of the conventional force fields. In the current study, the atomic Lennard-Jones parameters are taken directly from the CHARMM¹¹ protein force field without modification, and pair parameters are obtained by the usual combining rules. Furthermore the NDDO parameters for nonboundary atoms are taken from Austin Model 1 (AM1⁴⁴) without modification. The semiem-

pirical parameters for the carbon boundary atom are the same as in AM1 except that the values of U_{ss} and U_{pp} are scaled by 0.99 as in previous studies.^{34,36}

In calculating E_{elec} , the electric potential due to fragments sharing a boundary atom with the QM fragment under current consideration is calculated by explicit Coulomb integrals;³⁶ the electric potential from non-neighboring fragments are approximated by one-electron integrals with partial atomic charges. In the present calculations these charges are obtained by the Mulliken approximation⁴⁵ applied to wave functions of each of the other fragments.

The total electronic energy of the system is determined by a double self-consistent-field (DSCF) procedure.^{32,34–36} Starting with an initial guess of the one-electron density matrix for each fragment, one cycles over all fragments in the system and performs electronic structure calculations for each fragment (peptide unit or water molecule) in the presence of Mulliken charges of the other fragments until the change in total electronic energy or density matrix satisfies a predefined tolerance.^{32,34,36} To facilitate the DSCF convergence, we introduce a quantum mechanical buffer zone for the peptide unit (m) currently being treated quantum mechanically in the inner SCF iteration.^{35,36} Thus, in addition to this fragment m , we also include the peptide units prior to and after fragment m in each explicit QM treatment. In turn, fragment m becomes a buffer fragment for peptide units $m - 1$ and $m + 1$, respectively. Note that during the SCF optimization of the wave function for fragment m , the electron densities of the buffer peptide units are kept frozen³⁶ at values derived from a previous outer SCF iteration. Although it increases the number of two-electron integrals, the use of a buffer zone³⁶ reduces the time spent on matrix transformations (as compared to the earlier formulations) because no atom needs special treatment to avoid double counting or unphysical interactions with virtual orbitals. Once the wave function is converged, the forces are calculated analytically.³⁵ When the DSCF process has converged, the chemical potentials of all fragments will have been equalized. This allows mutual polarization of all fragments subject to the constraint that there is no charge transfer between fragments.

A key methodological issue is that the Fock matrix is expressed in a mixed basis consisting of atomic orbitals for nonboundary atoms and hybrid orbitals^{34,39} for boundary atoms. The present usage of hybrid orbitals is expected to be more accurate than the original use³⁹ because the charges in the hybrid orbitals are all determined self-consistently rather than determining some of them from MM parameters.

References 32 and 34–36 contain further details of the method.

3. Computational Details

The initial structure of a BPTI protein molecule solvated in a cubic box of water molecules is constructed using a developmental version of CHARMM (version c34a1),⁴⁶ in which the present X-Pol potential has been implemented. The non-hydrogen atomic coordinates are taken from the structure 6PTI in the protein databank (PDB), and all hydrogen atoms are built using the HBUILD function in

CHARMM based on standard equilibrium geometrical parameters.¹¹ There are two disulfide bonds in BPTI, which are terminated by hydrogen atoms in the X-Pol treatment to allow for convenient partition of the protein into peptide units. We note that a simple extension of the procedure already implemented into CHARMM can be made to handle disulfide bond connecting two amino acids. For the present study, our current treatment is sufficient. We have used a neutral side chain for each histidine residue, while all other titratable residues are assigned a protonation state corresponding to a pH of 7. The BPTI protein is then solvated by a previously equilibrated water box of about $54 \times 54 \times 54 \text{ \AA}^3$, deleting all water molecules within 2.7 \AA of any protein atoms, resulting in a total of 4461 water molecules, giving rise to a total of 14281 atoms, including protein, solvent, and counterions. There are 4519 fragments (58 amino acid residues and 4461 waters).

An MD simulation in the *NPT* ensemble at 300 K and 1 atm is carried out for 100 ps using the CHARMM22 force field¹¹ for protein and the three-point-charge TIP3P model⁴⁷ for water to equilibrate the system. The resulting coordinates are used as the initial structure for a 50 ps *NVT* MD simulation at 300 K using the X-Pol potential.^{34–36} The box length is set to 53.65 \AA in the X-Pol calculations, which is the average value from the MD simulation using the CHARMM22 force field. A Nosé-Hoover thermostat⁴⁸ is used for temperature control.

All simulations utilized an integration time-step of 1 fs, and the SHAKE algorithm⁴⁹ is used to constrain bond distances involving hydrogen atoms at the equilibrium values defined in the CHARMM22 force field.¹¹ Electrostatic interactions between fragment pairs whose centers of mass are separated beyond 11 \AA are truncated (shifting or switching can be introduced as a refinement in later work). The convergence criterion for average diagonal elements of the density matrix is set to 10^{-6} .

All calculations were carried out using a locally modified version of the CHARMM⁴⁶ software package. The X-Pol potential was initially developed based on a new semiempirical code written by Walker et al.,⁵⁰ and the current X-Pol software is essentially entirely rewritten.

4. Results and Discussion

A snapshot of the BPTI structure at the end of the 50 ps MD simulation using the X-POL potential is displayed in Figure 2 along with the structure at the end of a 50 ps MD simulation using the CHARMM22 force field and the crystal structure 6PTI from the protein data bank (PDB). These figures show that the secondary structures of BPTI retained their configurations in 50 ps molecular dynamics simulations employing the X-Pol potential (Figure 2a) in comparison with the crystal structure (Figure 2c). It appears that the two β -strand configurations are somewhat weakened from simulations using the classical force field. Compared with the crystal structure, the side chains show significant conformational change from both classical and X-Pol molecular dynamics simulations; charged residues are more exposed to the solvent on the protein surface. One realizes that the semiempirical AM1 model was not developed to treat

intermolecular interactions accurately. Thus, we note that the present X-Pol potential with the original AM1 quantum mechanical method is not a quantitatively accurate force field for performing quantitative studies of the dynamics of solvated proteins. To achieve this goal, a reparameterized and well-tested QM model is needed, and the development of such a model is left for future research.

The fluctuation of the total potential energy is displayed in Figure 3 for the entire 50 ps (50000 steps) of simulation, showing that the energy exhibited an initial rise in the first 15 ps and then quickly settled to a stable average throughout the rest of the simulation. At each MD step, about 7 iterations were sufficient to achieve SCF convergence to an accuracy of 10^{-6} in the electronic one-electron density matrix.

The main physical result of the present study is the extent of electronic polarization and intramolecular charge transfer in the solvated protein. The net charge from Mulliken population analysis of the wave function for each carbonyl group (C=O) in the protein backbone is calculated and averaged for the last 30000 MD steps. The average net charge on the backbone carbonyl (C=O) group of each residue along the peptide chain is shown in Figure 4. We found that all carbonyl groups bear a negative net charge, which is reasonable since C=O is a strong electron withdrawing group. The average net charges on the carbonyl groups range from -0.05 to -0.16 (all partial charges are in units of a proton charge), with 17 of them more negative than -0.10 . In comparison, the CHARMM22 force field¹¹ employs fixed partial atomic charges with the convention that the net group charge for each carbonyl unit (C=O) is zero in the protein backbone. Since it is computationally efficient for each group charge to be zero (forcing groups of 2–10 atoms to be perfectly neutral allows not only for more easily transferable charge parameters but also for more efficient truncation of long-range electrostatics^{3a}), this can only be remedied in conventional molecular mechanics calculations by using larger units as groups and by parametrizing the groups to allow different charges on carbonyl groups in different environments. However, even if that is done, the charge on each carbonyl group would be independent of time and environment, neither of which is found to be the case in the X-Pol calculations.

To analyze charge transfer effects between different residues, we calculated the net charge of each residue by Mulliken population analysis. Here, we note that formally there is no charge transfer between fragments treated in the X-Pol potential. However, effective charge transfer can be observed through the boundary atom due to intrafragment electronic polarization. Thus, it is possible that the electron density of the two active orbitals in the (*I*-1)*th* residue (green fragment in Figure 1) is depleted into the rest of the fragment, whereas the two active orbitals in the *I**th* residue (red in Figure 1) attract greater charge density in that fragment. Thus, the net partial atomic charge on the boundary atom has contributions from the charge densities of both neighboring residues. In the following discussion, the term “charge transfer” or excess charge is used to describe the difference of the total Mulliken population charge of a formal residue (not the peptide unit used in the definition of the QM

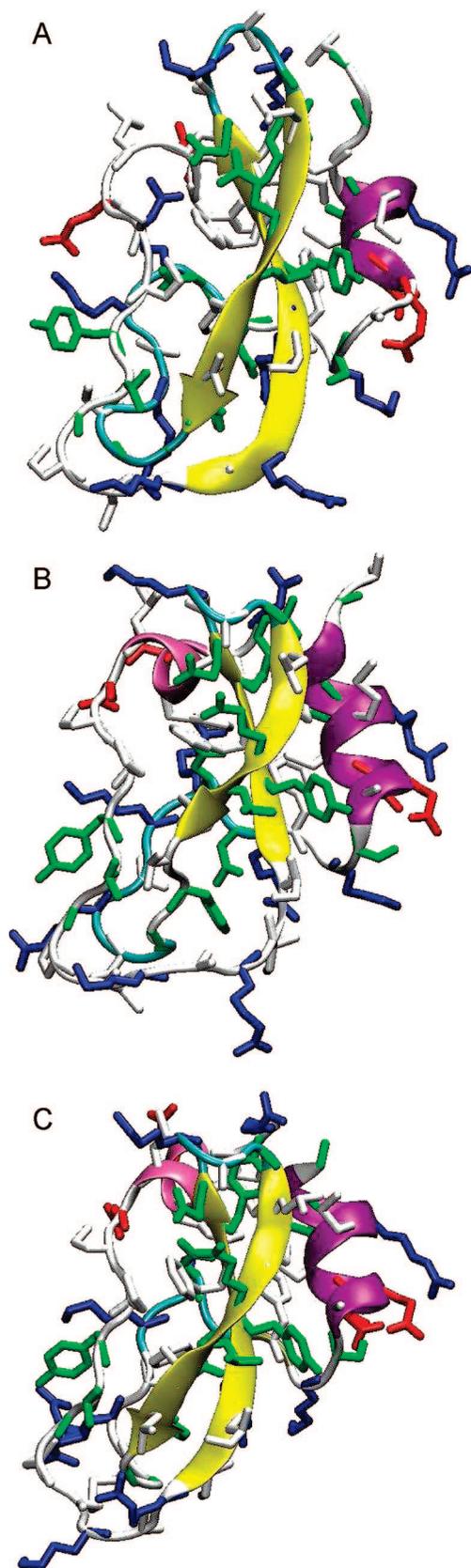


Figure 2. (a) Snapshot of the BPTI structure from MD simulation with the X-Pol potential, (b) snapshot of the BPTI structure from MD simulation with the CHARMM22 force field, and (c) X-ray crystal structure. Secondary structures are shown in yellow for β -strands, in purple for α -helices, and in gray and cyan for loops. Side chains are depicted in gray for hydrophobic residues, in green for polar residues, in blue for cationic residues, and in red for anionic residues.

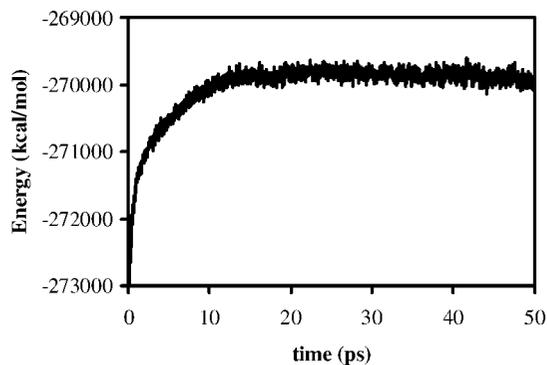


Figure 3. Histogram of the potential energy (kcal/mol) during the 50 ps molecular dynamics simulation of BPTI in water using the X-Pol potential.

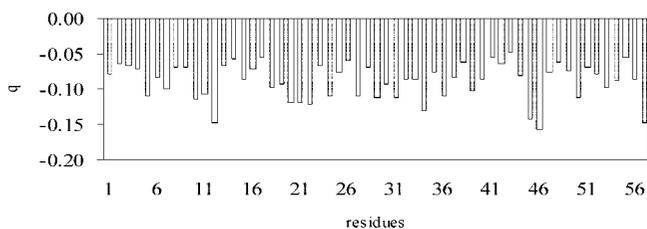


Figure 4. Average net partial charges (in atomic units) on the backbone carbonyl (C=O) group of each amino acid residue in BPTI. The carbonyls are arranged in order of sequence number.

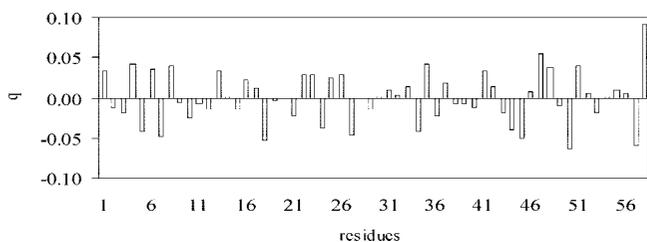


Figure 5. Average excess charge (in atomic units) for each residue of BPTI in water. The residues are arranged in order of sequence number. Note that Figures 5, 6, and 7(b) refer to conventional residues, not to peptide units.

fragment) in the protein and that of an isolated residue. Thus, the dominant contributor to charge transfer in this model is intrafragment polarization in the X-Pol representation. If one is interested in charge transfer between two residues, for example, in an ion pair, salt bridge, or hydrogen bond to a charged residue, the two moieties between which charge transfer is to be allowed should be treated as a single fragment.

Figure 5 shows the excess charge (calculated by subtracting the formal charge associated with the protonation state from the Mulliken charge) for all residues averaged over the last 30000 steps of MD simulations. The average excess charge ranges from -0.06 to $+0.09$. The excess charge of the same type of amino acid at different locations in the protein is displayed in Figure 6. This figure shows that charge transfer can be quite different depending on the specific position of a given residue as well as its environment and the protein sequence, for example, the average excess charge on phenylalanine ranges from -0.05 to $+0.04$, that on cysteine from -0.04 to $+0.04$, and that on alanine from

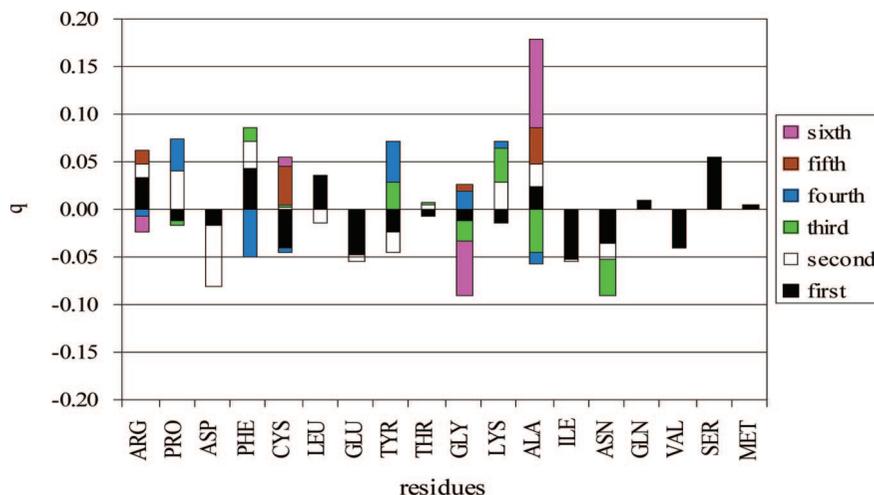


Figure 6. Average excess charge (in atomic units) for each residue of BPTI in water. The results for peptide units of the same type are shown together. The ordering of the residues, to which the colors refer, is by sequence number. Negative excess charges are shown below the baseline, and positive charges above it. The magnitude of the charge associated with each colored segment is from the bottom of a given colored segment to the top, not from the baseline to the top/bottom of each color bar. So the magnitude of charge associated with the residue with the longest colored segment is the largest.

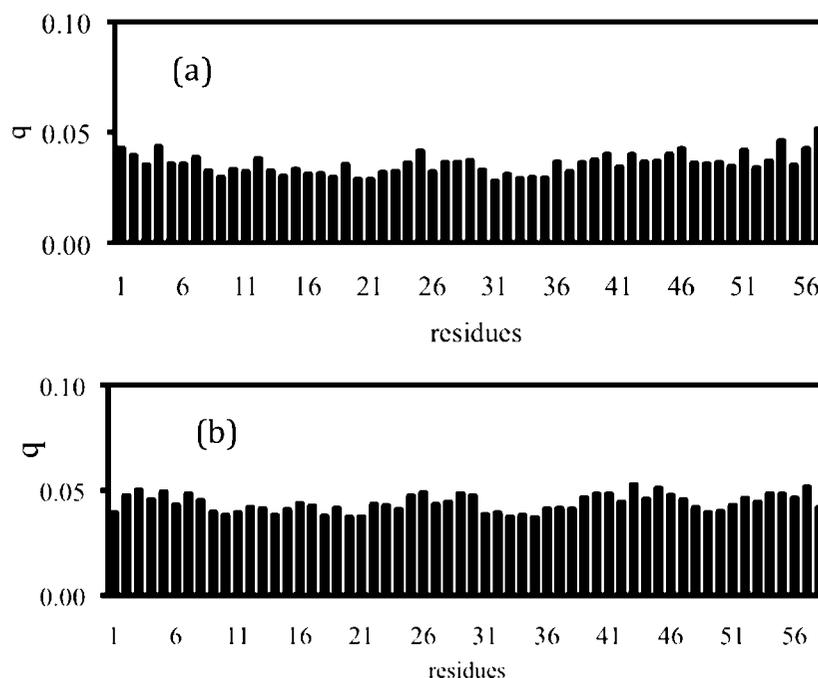


Figure 7. Standard deviations of the charges shown in Figures 4 and 5: (a) partial charges (in atomic units) on the backbone carbonyls and (b) excess charges on the residues.

−0.05 to +0.09. The instantaneous excess charges vary even more widely.

In Figures 4–6 we reported averages for net carbonyl charges and for excess residue charges over 30000 time frames. Using the same data, we also computed standard deviations for these 30000 time frames, and this provides a measure of the variation of the instantaneous charges with time (as the environment changes). These results are given in Figure 7. This variation is different from the sequence dependence of the average (shown in Figure 6), and it is a feature of the true dynamics that a nonpolarizable force field can never reproduce. Figure 7 shows that the standard deviations are in the range of 0.03 to 0.05.

5. Computational Considerations and Future Improvements

The total CPU time for the 50 ps simulation using a single 2.66-GHz SGI Altix XE 1300 Linux Cluster processor is 377 h, whereas it took 62.6 h to run 5 ps on a single 1.5-GHz IBM Power4 processor. During each DSCF calculation, all one- and two-electron integrals are saved in the memory although one has the option to calculate the electron integrals on the fly. The number of one-electron QM/MM integrals scales as N^2 where N is the number of fragments, and the total memory used to store one-electron integrals is about 458 MB; however, the two-electron integrals are only evaluated within each fragment and its

buffer fragments that share a boundary atom, so their number scales as N , and the total memory used to store two-electron integrals is merely 13 MB. In principle, by reorganizing the algorithm, the X-Pol potential is highly parallelizable, at least up to a number of processors equal to the number of fragments, since fragments or groups of fragments and the associated QM/MM integrals can be distributed over processors, and the only messages to be communicated for potential energy evaluation are the DSCF updates to auxiliary density matrices, the buffer density matrices, and the partial charges. A parallel version of the X-Pol potential is currently being developed.

In the present X-Pol potential, atom-centered point charges are calculated from class II Mulliken population analysis. In future work, it may be worthwhile to explore other approximations to the MM potential such as charges fitted to electrostatic potentials (class III ESP charges),⁵¹ CM4 class IV charges,⁵² or distributed multipoles.⁵³

In the present study, charge transfer along the chain occurs through boundary atoms, whereas charge transfer between protein and solvent is not included. Although the present study used small fragments (peptide units and water molecules), the method is very flexible, and one can use larger fragments to include charge transfer between fragments not connected by a chain of bonds or to include charge transfer between connected fragments more self-consistently. For an example of the latter, one can include in a single fragment either neighboring peptide units or peptide units interacting in secondary or tertiary structure via hydrogen bonds or salt bridges. As mentioned above, one can include solute-solvent charge transfer by, for example, treating a peptide unit and a nearby water molecule (or molecules) as a single fragment. Since water molecules exchange in and out of the first solvation shell, such solute-solvent fragments should be treated by an adaptive⁵⁴ algorithm that allows such exchanges. For studying enzyme kinetics, one could take the entire substrate and coenzyme to be a single fragment, if desired.

An important objective in future work is to carefully parametrize the quantum mechanical model to achieve the desired accuracy for properties both in the gas phase and in the condensed phase.

6. Concluding Remarks

We have demonstrated the applicability of an explicit polarization (X-Pol) method³¹⁻³⁶ to a solvated protein in a water box with $>10^4$ atoms. The calculation presented here is equivalent to a molecular dynamics simulation of a 14281-atom system consisting of 29026 basis functions with direct dynamics based on an explicit quantum mechanical electronic wave function for the entire system plus a van der Waals term for interfragment exchange repulsion and dispersion forces. The present molecular dynamics simulation generated a trajectory over a 50 ps time interval in an NVT ensemble using an existing semiempirical model. Remarkably, on a single processor without carefully optimizing the quantum mechanical code, it is possible to run more than 3 ps of direct dynamics per day. Whereas all atoms are treated quantum mechanically in this simulation, a typical combined QM/

MM simulation treats only 10^1-10^2 atoms quantum mechanically (for example, the quantum-electronic-structure subsystem had 69,⁵⁵ 47-60,⁵⁶ 53-56,⁵⁷ 102,⁵⁸ and 71⁵⁰ atoms in some recent studies). Analysis of the wave function implies that the polarization and charge transfer effects are significant in the condensed phase and protein. Water molecules display a significant polarization effect in the condensed phase. Carbonyl groups in the protein backbone bear a negative net charge. The net charges of the backbone carbonyl group in different amino acid residues are different by an amount up to about 0.1 e which suggests that in some cases intramolecular charge transfer needs to be considered explicitly. Analysis of the excess charge of each amino acid shows fluctuating charge transfer between amino acid residues in the protein. The same residue may have significantly different charge transfer effects depending on protein sequence.

In closing, it is worthwhile to emphasize the differences between the present linear scaling method and problem decomposition by a divide-and-conquer-type⁵⁹ approach. Methodologically, X-Pol and divide-and-conquer are not the same, even when the same quantum mechanical model is used. Divide-and-conquer is a linear scaling method to efficiently obtain a solution of the quantum mechanical model (e.g., Hartree-Fock or Kohn-Sham equations) for a large system. In contrast, X-Pol is a quantum mechanical force field, whose energy is not the Hartree-Fock or Kohn-Sham energy of the system. X-Pol is variational and constitutes an efficient method that can be used to run molecular dynamics simulations for a fully solvated protein, whereas D&C requires many more SCF iterations to obtain an energy that is accurate enough for gradients (forces) in MD simulations. Although AM1 was used in the present study for the purpose of proof of concept, the X-Pol potential is designed as a force field to be parametrized just as "standard" CHARMM, AMBER, or OPLS force fields are parametrized.

Acknowledgment. This work was supported in part by the National Institutes of Health under award no. GM46736, the Office of Naval Research under grant no. N00012-05-01-0538, and the National Science Foundation under grant no. CHE07-04974. J.G. thanks the Ministerio de Educación y Ciencia, España for partial support during his sabbatical leave and the BSC for human and technical resources.

References

- (1) (a) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186. (b) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151. (c) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679. (d) Rueda, M.; Ferrer, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796.
- (2) Bixon, M.; Lifson, S. *Tetrahedron* **1967**, *23*, 769.
- (3) (a) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27. (b) Patel, S.; Brooks, C. L., III. *Mol. Simul.* **2005**, *32*, 231. (c) Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182. (d) MacKerell, A. D., Jr. *Annu. Rep. Comput. Chem.* **2005**, *1*, 91.

- (4) (a) Xie, W.; Pu, J.; MacKerell, A. D., Jr.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1878. For a thematic issue devoted to current research on polarization and polarizable force fields, see: (b) Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.
- (5) (a) Hill, T. *J. Chem. Phys.* **1946**, *14*, 465. (b) Westheimer, F. H.; Mayer, J. E. *J. Chem. Phys.* **1946**, *14*, 733.
- (6) Hendrickson, J. B. *J. Am. Chem. Soc.* **1961**, *83*, 4537.
- (7) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
- (8) For a historical account of the development of the Lifson-style force field, see Levitt, M. *Nat. Struct. Mol. Biol.* **2001**, *8*, 392.
- (9) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.
- (10) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (11) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (12) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (13) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Eur. Biophys. J.* **2005**, *34*, 273.
- (14) Maple, J. R.; Hwang, M. J.; Jalkanen, K. J.; Stockfisch, T. P.; Hagler, A. T. *J. Comput. Chem.* **1998**, *19*, 430.
- (15) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.
- (16) Van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. J. *J. Mol. Biol.* **1987**, *198*, 721.
- (17) Niesar, U.; Corongiu, G.; Clementi, E.; Kneller, G. R.; Bhattacharya, D. K. *J. Phys. Chem.* **1990**, *94*, 7949.
- (18) Sprik, M.; Klein, M. L.; Watanabe, K. *J. Phys. Chem.* **1990**, *94*, 6483.
- (19) Dang, L. X.; Chang, T. M. *J. Chem. Phys.* **1997**, *106*, 8149.
- (20) (a) Rick, S.; Stuart, S.; Berne, B. *J. Chem. Phys.* **1994**, *101*, 6141. (b) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515. (c) Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. *J. Chem. Theory. Comput.* **2005**, *1*, 694.
- (21) Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1995**, *99*, 11575.
- (22) (a) Gao, J.; Habibollahzadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460. (b) Gao, J.; Pavelites, J. J.; Habibollahzadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689.
- (23) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.
- (24) (a) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933. (b) Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671. (c) Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quantum Chem.* **2007**, *107*, 1390.
- (25) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.
- (26) Cubero, E.; Luque, F. J.; Orozco, M.; Gao, J. *J. Phys. Chem. B* **2003**, *107*, 1664.
- (27) (a) Rappé, A.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358. (b) Ogawa, T.; Kuita, N.; Sekino, H.; Kitao, O.; Tanaka, S. *Chem. Phys. Lett.* **2004**, *397*, 382.
- (28) Vorobyov, I. V.; Anisimov, V. M., Jr. *J. Phys. Chem. B* **2005**, *109*, 18988.
- (29) Patel, S.; Brooks, C. L., III. *J. Chem. Phys.* **2005**, *123*, 164502.
- (30) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Dehez, F.; Angyan, J. G.; Orozco, M.; Chipot, C.; Luque, J. *J. Chem. Theory Comput.* **2007**, *3*, 1901.
- (31) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.
- (32) Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346.
- (33) Wierzychowski, S. J.; Kofke, D. A.; Gao, J. *J. Chem. Phys.* **2003**, *119*, 7365.
- (34) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890.
- (35) Xie, W.; Song, L.; Truhlar, D.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.
- (36) Xie, W.; Song, L.; Truhlar, D.; Gao, J. *J. Phys. Chem. B* **2008**, *112*, 14124.
- (37) Pulay, P. *Adv. Chem. Phys.* **1987**, *69*, 241.
- (38) *Pure Appl. Chem.* **1974** *40* 291 See also: Moss, G. P. *Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains*, IUPAC-IUB Commission on Biochemical Nomenclature (CBN): London, 1974. Available at <http://www.chem.qmul.ac.uk/iupac/misc/pppep1.html> (accessed Dec 20, 2008).
- (39) (a) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714. (b) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336.
- (40) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
- (41) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
- (42) (a) Gao, J.; Xia, X. *Science* **1992**, *258*, 631. (b) Gao, J. *Rev. Comput. Chem.* **1995**, *7*, 119.
- (43) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.
- (44) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- (45) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
- (46) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (48) (a) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511. (b) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (49) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (50) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.
- (51) (a) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592. (b) Gao, J.; Luque, F. J.; Orozco, M. *J. Chem. Phys.* **1993**, *98*, 2975.
- (52) (a) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87. (b) Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**,

- 100, 16385. (c) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820. (d) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
- (53) (a) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325. (b) Chipot, C.; Ángyán, J. G. *New J. Chem.* **2005**, *29*, 411.
- (54) (a) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231. (b) Heyden, A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 217.
- (55) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.
- (56) Spiegel, K.; Rothlisberger, U.; Carloni, P. *J. Phys. Chem. B* **2004**, *108*, 2699.
- (57) Rodriguez, A.; Oliva, C.; Gonzalez, M.; Van der Kamp, M.; Mulholland, A. J. *J. Phys. Chem. B* **2007**, *111*, 12909.
- (58) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159.
- (59) (a) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438. (b) Stechel, E. B. In *Domain-Based Parallelism and Problem Decomposition Methods in Computational Science and Engineering*; Keyes, D. E., Saad, Y., Truhlar, D. G., Eds.; SIAM: Philadelphia, 1995; pp 217–238. (c) Bowler, D. R.; Aoki, M.; Goringe, C. M.; Horsfield, A. P.; Pettifor, D. G. *Modell. Simul. Mater. Sci. Eng.* **1997**, *5*, 199. (d) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085. (e) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. *J. Chem. Phys.* **2000**, *113*, 5604. (f) Shimojo, F.; Kalia, R. K.; Nakano, A.; Vashishta, P. *Comput. Phys. Commun.* **2005**, *167*, 151.

CT800239Q