Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics

ARTICLE *in* JOURNAL OF PHYSICAL CHEMISTRY LETTERS · JULY 2011

Impact Factor: 7.46 · DOI: 10.1021/jz200697c

3 AUTHORS, INCLUDING:

Todd J. Martinez
Stanford University
239 PUBLICATIONS **7,875** CITATIONS

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

LETTER

pubs.acs.org/JPCL

Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics

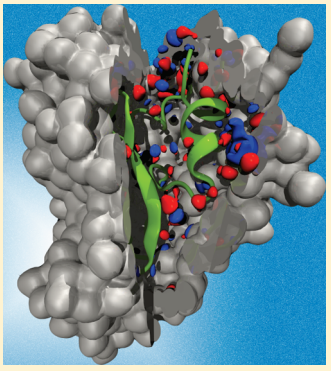
Ivan S. Ufimtsev,[†] Nathan Luehr,[†] and Todd J. Martinez*[,†,§]

[†]Stanford University, Stanford, California 94305, United States

[§]SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States

Supporting Information

ABSTRACT: Charge transfer at the Bovine pancreatic trypsin inhibitor (BPTI) protein−water interface was analyzed by means of ab initio Born−Oppenheimer molecular dynamics simulation of the entire protein running on graphical processing units (GPUs). The efficiency of the GPU-based quantum chemistry algorithms implemented in our TeraChem program enables us to perform these calculations on a desktop computer. Mulliken and Voronoi deformation density (VDD) population analysis reveals that between 2.0 and 3.5 electrons are transferred from surrounding water molecules to the protein over the course of the 8.8 ps simulation. Solving for the electronic structure of BPTI in the absence of surrounding water molecules (i.e., in the gas phase) leads to large intraprotein charge transfer, where approximately one electron in total is transferred from neutral to polar residues. Solvation relieves this polarization stress, leading to a neutralization of the excess positive charge of the neutral residues.



SECTION: Biophysical Chemistry

It has been long recognized that water plays an important role in protein structure and dynamics. Water on the protein surface, often referred to as biological water,[1] is an essential element of protein interactions[2] and enzyme function.[3] Some water molecules reside in the same location near the protein surface for a long time[1] compared with the typical relaxation time under bulk conditions. These water molecules form strong hydrogen bonds[4] and can be directly observed in accurate model-free crystallographic experiments.[5] Classical force fields have made tremendous progress in describing interactions at protein−water interfaces and can accurately predict such important energetic properties as solvation free energies of amino acids.[6,7] However, most of these theoretical models use a simplified "charged ball-and-spring" representation that is incapable of describing quantum mechanical phenomena like charge transfer (CT) and electronic polarization. Recently, it was demonstrated that CT effects account for approximately one-third of the binding energy in a neutral water dimer,[8] and for stronger H-bonds, one can anticipate this contribution to be even larger. Although CT and polarization effects are typically parametrized in classical force fields implicitly as a part of the electrostatic and Lennard-Jones two-body interactions, it remains an open question as to how accurately such approximations can describe biological water. Another recent study has stressed the importance of CT interactions in proteins and suggested this missing term should be explicitly included in future classical force field parametrizations.[9]

Although the effect of explicit solvent on protein structure and function has been studied for more than two decades,[4] solvated proteins have almost exclusively been treated using nonpolarizable classical force fields. Only a few attempts have been made to study protein−water systems at higher levels of theory, such as semiempirical[10−12,35] or fragment molecular orbital[13] approaches. However, even these efforts have still relied on molecular dynamics (MD) simulations with classical force fields to provide atomic coordinates for higher level calculations. More rigorous treatment of solvated proteins by means of Hartree−Fock (HF) or density functional theory (DFT) methods is clearly needed. Ideally, one would use ab initio rather than classical MD trajectories in such calculations because classical and ab initio dynamics could potentially sample configurational space quite differently. In fact, DFT MD has been applied to study model systems such as solvated glycine dipeptide,[14] and substantial CT was observed in these simulations. However, to the best of our knowledge, ab initio (HF or DFT) MD has never been used to treat entire proteins. The major obstacle to the use of ab initio methods in this context is their high computational cost. Recent single-point energy calculations of solvated rubredoxin represent an illustrative example.[15] Calculation of the energy for the resulting 2825 atoms required over 1 h on 8196 processor cores. Dynamical simulations requiring hundreds or thousands of such calculations would appear to be completely out of reach. Fortunately, graphical processing units (GPUs) (essentially consumer videogame graphics cards) have emerged as a powerful alternative to traditional processors. We have redesigned algorithms for electronic structure theory and ab initio MD to leverage the strengths of GPUs, with promising results.[16−18] In this Letter, we

Received: May 24, 2011
Accepted: June 27, 2011

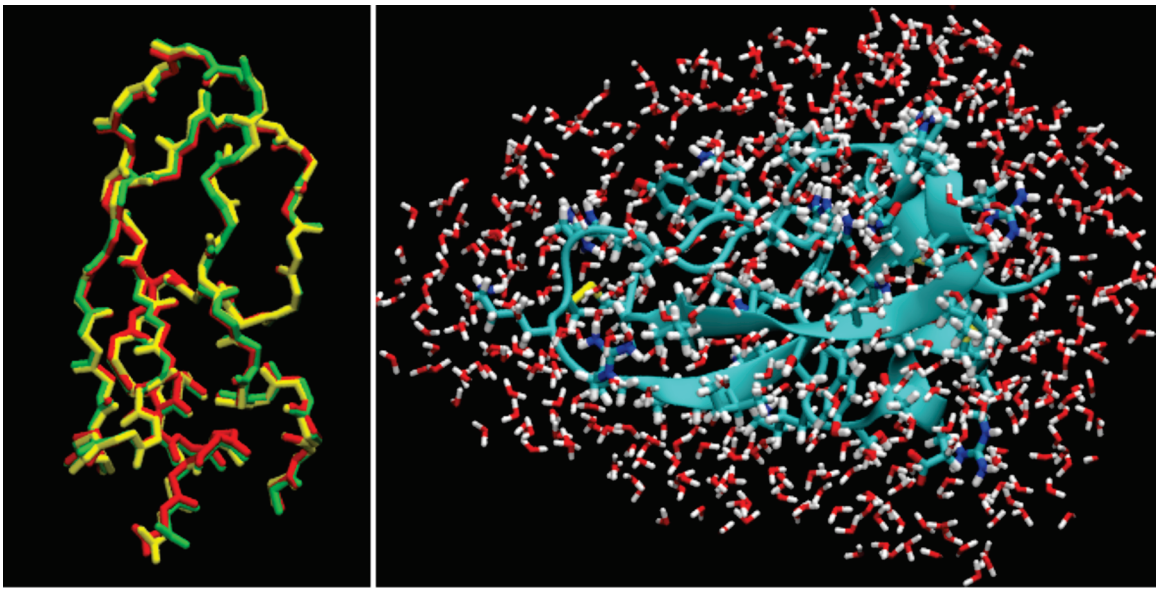


Figure 1. Left: overlaid BPTI crystal structure (red) and the same structure optimized at RHF/STO-3G (yellow) and RHF/6-31G (green) levels. Right: solvated BPTI protein used in the charge transfer analysis. The system contains 2634 atoms and 12527 basis functions when treated at RHF/6-31G and B3LYP/6-31G levels.

exploit these advances[19] to study CT at the protein−water interface. The potential for CT has profound implications for simulations using classical force fields because CT effects cannot be described at all without at least polarizable force fields,[20−22] yet this effect can be an important component to the stability of solvated proteins.[10,11] To generate a set of protein−water structures for CT analysis, we perform the first ab initio Born−Oppenheimer MD simulation of an entire protein, bovine pancreatic trypsin inhibitor (BPTI). Coincidentally, this is also the first protein ever studied by MD simulation with classical force fields.[23,24] All calculations were performed by TeraChem,[19] a general purpose quantum chemistry package designed from scratch for GPU architectures.

First, we assessed the ability of ab initio "force fields" to retain the crystal structure of BPTI by carrying out geometry optimization of the entire protein structure taken from the Protein Databank (accession code: 6PTI) including all crystallographic waters and the phosphate group. The electronic structure was solved by the restricted Hartree−Fock (RHF) method with the STO-3G and 6-31G basis sets. Both optimized structures remain quite close to the reference crystal structure with root-mean-squared deviation (rmsd) of aligned backbone atoms of 0.42 and 0.31 Å for RHF/STO-3G and RHF/6-31G, respectively. The left panel in Figure 1 depicts the overlaid backbones of the experimental (red) and optimized STO-3G (yellow) and 6-31G (green) structures.

Proteins exhibit high conformational flexibility, which is an essential part of their function. Therefore, CT and other properties are dynamical quantities, which can be strongly influenced by protein−water conformational dynamics. To sample the protein configurational space, we performed an 8.8 ps ab initio MD simulation of BPTI solvated by 578 water molecules at constant temperature (T = 300 K). The entire protein along with six crystallographic water molecules located inside the two protein pockets (900 atoms in total) was treated at the RHF/STO-3G level, whereas the exterior solvent molecules were described by the classical TIP3P force field, that is, a QM/MM simulation[25,26] with the entire protein in the QM region. The system had a total charge of +6e, consistent with experimental measurements for BPTI in gas phase and solution.[27,28]

Figure 2 depicts structural changes of the protein during the MD simulation in terms of the rmsd of BPTI backbone atoms from the reference protein structure that represents the starting geometry at zero time (black line in Figure 2). We have also generated a short classical MD trajectory of the same system using the NAMD[29] package and CHARMM31[30] classical force field for comparison (blue line in Figure 2). Both simulations start from the same experimental crystal structure of BPTI. Although the system was not pre-equilibrated, one can see in Figure 2 that the ab initio rmsd curve shows signs of stabilization around 1.5 Å. The protein structure was preserved, and at least a partial equilibrium state was reached during the ab initio MD simulation. Furthermore, the classical and ab initio curves follow each other rather closely, indicating similar structural evolution of the systems. Curiously, the ab initio MD simulation remains a little closer to the starting crystal structure (as evidenced by the lower rmsd throughout the simulation).

Eighty-eight geometries separated by 100 fs time intervals were extracted from the QM/MM MD trajectory for further analysis. One of these structures is depicted in the right panel of Figure 1. The magnitude of protein−water CT in each structure was quantified by single-point energy calculations at RHF/6-31G and B3LYP/6-31G levels, followed by Mulliken population analysis (PA) for (1) a QM/MM calculation with QM BPTI and MM explicit water and (2) a fully QM calculation of the entire BPTI−water system (as in Figure 1, right). In the latter case, this amounts to 2634 atoms and 12 527 basis functions, which until now would tax even a supercomputer. Using a single desktop workstation with eight GPUs operating in parallel, we were able to solve each of the 88 electronic structure problems in <3 h. For every atom in the protein, we calculated the change of its partial charge $\Delta q_i = q_i^{\mathrm{BPTI(QM)+solvent(MM)}} - q_i^{\mathrm{BPTI(QM)+solvent(QM)}}$ induced by allowing for protein−water CT. These solvent-induced charge differences were then summed for all neutral, polar, and

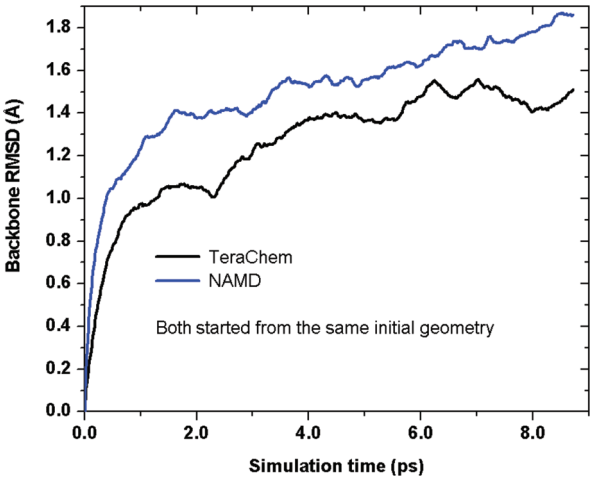


Figure 2. Comparison of time evolution of rmsd from the experimental BPTI structure for empirical force field dynamics (CHARMM31 force field using NAMD, blue) and ab initio molecular dynamics using TeraChem (black). Both simulations are at 300 K and start from the experimentally determined crystal structure.

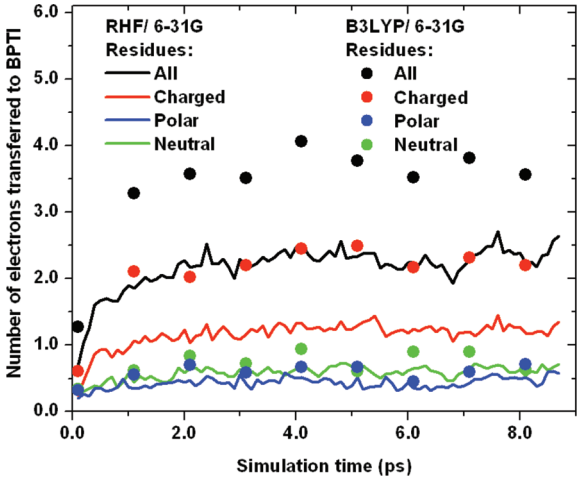


Figure 3. Number of electrons transferred from water to BPTI. The total transferred charge (black) is a sum of individual contributions from charged (red), polar (blue), and neutral (green) residues. The calculations were carried out at the RHF/6-31G (solid curve) and B3LYP/6-31G (circles) levels of theory. Approximately 2.6 and 3.6 electrons are transferred by the end of the RHF and DFT simulations, respectively.

charged residues separately to evaluate the individual contribution of different types of amino acids to the overall CT effect. Proline and cysteine amino acids were counted as polar residues for the purpose of this analysis. We also performed analogous calculations on 22 geometries extracted from the same MD trajectory using the 6-31G* basis set, and the result was essentially the same as that obtained using the 6-31G basis set, with the largest deviation being <5%.

The results of the CT analysis are presented in Figure 3, where the black curve depicts the total amount of charge transferred from BPTI to the solvent (or, equivalently, the number of electrons transferred from the solvent to BPTI), which is the sum of all induced Δq_i in the protein. In addition, individual contributions from charged, polar, and neutral residues are represented in red, blue, and green, respectively. Approximately

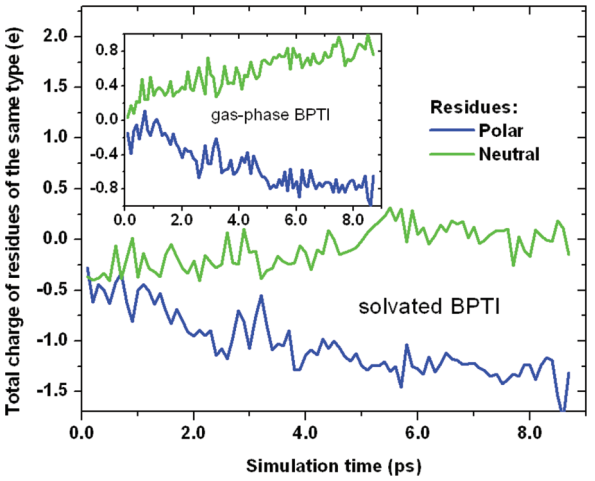


Figure 4. Total charge of polar (blue) and neutral (green) residues of solvated BPTI protein. Inset: same for the isolated protein. In gas phase, neutral amino acids donate electrons to polar residues (approximately one electron in total). This effect is not observed for the solvated protein, where the neutral amino acids remain nearly neutral.

2.6 electrons are transferred to the protein by the end of the simulation, making the net BPTI charge deviate significantly from the classical +6*e* value typically assigned based on standard pH rules. This effect is even larger (3.6 to 4.0 electrons, dotted lines in Figure 3) if DFT is used to solve the electronic structure problem. This is likely at least partially due to the self-interaction error and the resulting tendency of DFT to delocalize electrons.[31] Furthermore, the time course of CT (and thus the total BPTI charge) reveals two kinds of behavior: rapid fluctuations caused by thermal motion and slower relaxation associated with conformational rearrangements inside the protein and at the protein−water interface. Comparison of the individual CT contributions with the net CT in Figure 3 shows that even though all types of amino acids participate in CT, the relaxation character of the net CT curve is mostly due to contribution of charged residues, or, in other words, due to solvent rearrangement around these particular residues as the system approaches equilibrium. The characteristic time of this process is ∼2 ps, which is in good agreement with that observed in spectroscopy experiments of water at protein surfaces.[1]

As expected, different types of amino acids participate in CT to different extents. For example, 15 charged residues (5 negatively and 10 positively charged) are involved in the transfer of 1.23 electrons in total, whereas 22 polar and 20 neutral residues are involved in the transfer of 0.49 and 0.64 electrons, respectively. We have averaged these numbers over the last 2 ps of the MD simulation. An interesting fact is that neutral amino acids transfer more charge than approximately the same number of polar residues. Note that CT was defined as a charge difference between QM/MM (rather than gas phase) and fully QM-solvated systems. Even though this definition at least partially excludes electronic polarization effects, the electronic density around neutral residues is nevertheless changed significantly upon solvation in QM water due to CT. Figure 4 depicts the total charge of polar (blue) and neutral (green) residues of solvated and gas-phase (inset in Figure 4) BPTI. One can see in Figure 4 that there is strong intraprotein CT (or polarization) in the gas-phase protein, where neutral residues donate approximately one electron to polar amino acids in total. This

dx.doi.org/10.1021/jz200697c |*J. Phys. Chem. Lett.* 2011, 2, 1789–1793

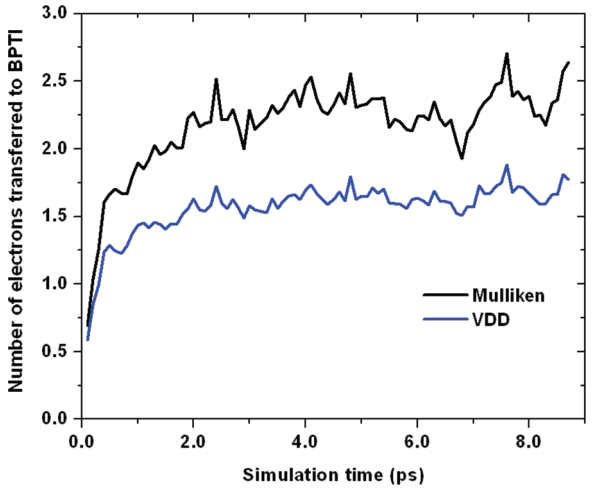


Figure 5. Total number of electrons transferred from water to BPTI, as calculated using Mulliken (black) and Voronoi deformation density (VDD, blue) population analysis. The two data sets not only demonstrate similar magnitudes of the charge transfer effect but also are strongly correlated with each other (R = 0.98). Thus, the Mulliken population analysis quantifies the coarse-grained charge transfer effect correctly.

imbalanced charge distribution could potentially interfere with protein function. Adding QM water completely neutralizes the excess positive charge of neutral residues. As shown in Figure S5 of the Supporting Information, QM/MM solvation is not sufficient to relieve this polarization stress. Therefore, polar and neutral amino acids not only effectively participate in CT at the protein−water interface but also play an active role in intraprotein CT. In contrast, charged amino acids do not contribute to intraprotein polarization in gas phase because their total charge fluctuates around +6.2e (in gas phase). The CT effect for charged amino acids is primarily determined by the solvent structure around these residues and is not notably affected by conformational rearrangements occurring inside the protein.

Mulliken population analysis (PA) has been criticized for providing unphysical atomic charges.[32,33] However, one can expect that such criticisms are less relevant when one calculates not atomic charges but rather total charges for large molecules and molecular fragments, as we do here (the entire protein and conglomerates of protein residues). This is because the major difficulty of orbital-based methods such as Mulliken PA is the nonorthogonality of the atomic orbitals and resulting basis set dependence of atomic charges. PA methods based on the electronic density are less sensitive to the choice of basis set and provide more robust atomic charges. Therefore, we have carried out the same CT analysis using the Voronoi density deformation[33] (VDD) scheme. Like other density-based methods, VDD is rather computationally intensive for large molecules because it involves evaluation of electronic density on a numerical grid. Therefore, similar to the electronic structure, we perform the VDD calculations on the GPU using a set of efficient GPU-accelerated routines that we developed for DFT calculations such as construction of numerical grids and evaluation of electronic density. The results are presented in Figure 5, which depicts the total number of electrons transferred to BPTI from water, as obtained from Mulliken (black) and VDD (blue) PA. Both curves show a similar number of transferred electrons (2.5 and 1.8 electrons for Mulliken and VDD PA, respectively) and are strongly correlated with each other (correlation coefficient, R = 0.98). Not only the total transferred charge but also the individual contributions of charged, polar, and neutral residues are strongly correlated (R = 0.98, 0.89, and 0.96, respectively). We therefore conclude that the observed CT effect is robust with respect to the choice of PA methods.

Water mediates protein function in many ways. It serves as a buffer for chemical reactions in enzymes, initiates hydrophobic collapse in protein folding, and actively participates in protein recognition. The results presented in this Letter demonstrate another potentially important role of biological water. Water molecules located near the protein surface notably affect the protein's electronic structure and stimulate release of polarization stress. This effect occurs in vacuo and leads to significant intraprotein CT. (See positively charged neutral residues in gas phase BPTI, as characterized in Figure 4.) Polarization stress could potentially lead to adverse consequences disrupting protein function. Upon solvation, however, BPTI restores its intraprotein charge balance by borrowing two to three electrons from the surrounding water bath, and the neutral residues acquire the expected, that is, neutral, charge. It is an open question whether CT effects like those observed here can be described by polarizable force fields.[34] Fortunately, the answer is now in reach because fast reference ab initio calculations are already possible, as we demonstrate. Our results reinforce previous findings obtained with more approximate methods for other proteins,[10−13] and one can anticipate that CT is an intrinsic characteristic of large biomolecules, essential for their function.

■ ASSOCIATED CONTENT

Supporting Information. Details of Mulliken charge analysis by residue for gas-phase, QM/MM-solvated, and QM/QM solvated structures at different times. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: todd.martinez@stanford.edu.

■ ACKNOWLEDGMENT

This work was supported by the National Science Foundation (CHE-06-26354) and the Department of Defense (Office of the Director of Defense Research and Engineering) through a National Security Science and Engineering Faculty Fellowship. I.S.U. is an NVIDIA Fellow.

■ REFERENCES

(1) Pal, S. K.; Peon, J.; Zewail, A. H. Biological Water at the Protein Surface: Dynamical Solvation Probed Directly with Femtosecond Resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1763–1768.

(2) Levy, Y.; Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.

(3) Zaks, A.; Klibanov, A. M. The Effect of Water on Enzyme Action in Organic Media. *J. Biol. Chem.* **1988**, *263*, 8017–8021.

(4) Levitt, M.; Sharon, R. Accurate Simulation of Protein Dynamics in Solution. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557–7561.

(5) Burling, F. T.; Weis, W. I.; Flaherty, K. M.; Brunger, A. T. Direct Observation of Protein Solvation and Discrete Disorder with Experimental Crystallographic Phases. *Science* **1996**, *271*, 72–77.

(6) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.
(7) Chang, J.; Lenhoff, A. M.; Sandler, S. I. Solvation Free Energy of Amino Acids and Side-Chain Analogues. *J. Phys. Chem. B* **2007**, *111*, 2098–2106.
(8) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Electron Donation in the Water-Water Hydrogen Bond. *Chem.—Eur. J.* **2009**, *15*, 851–855.
(9) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. n→π^* interactions in proteins. *Nat. Chem. Biol.* **2010**, *6*, 615–620.
(10) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M. Charge-Transfer Interactions in Macromolecular Systems: A New View of the Protein/Water interface. *J. Am. Chem. Soc.* **1998**, *120*, 5593–5594.
(11) Van der Vaart, A.; Merz, K. M. The Role of Polarization and Charge Transfer in the Solvation of Biomolecules. *J. Am. Chem. Soc.* **1999**, *121*, 9182–9190.
(12) Khandogin, J.; York, D. M. Quantum Descriptors for Biological Macromolecules from Linear-Scaling Electronic Structure Methods. *Proteins* **2004**, *56*, 724–737.
(13) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. Change in a Protein's Electronic Structure Induced by an Explicit Solvent: An Ab Initio Fragment Molecular Orbital Study of Ubiquitin. *J. Comput. Chem.* **2007**, *28*, 1750–1762.
(14) Dal Peraro, M.; Raugei, S.; Carloni, P.; Klein, M. L. Solute-Solvent Charge Transfer in Aqueous Solution. *ChemPhysChem* **2005**, *6*, 1715–1718.
(15) Guidon, M.; Hutter, J.; VandeVondele, J. Robust Periodic Hartree-Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets. *J. Chem. Theory Comput.* **2009**, *5*, 3010–3021.
(16) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222–231.
(17) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput* **2009**, *5*, 1004–1015.
(18) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619–2628.
(19) PetaChem. http://www.petachem.com (accessed 6/1/11).
(20) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
(21) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.
(22) Yu, H. B.; van Gunsteren, W. F. Accounting for Polarization in Molecular Simulation. *Comput. Phys. Commun.* **2005**, *172*, 69–85.
(23) Levitt, M.; Warshel, A. Computer Simulations of Protein Folding. *Nature* **1975**, *253*, 694–698.
(24) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585–590.
(25) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions. *J. Mol. Biol.* **1976**, *103*, 227–249.
(26) Friesner, R. A.; Guallar, V. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.
(27) Jarrold, M. F. Peptide and Proteins in the Vapor Phase. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179–207.
(28) Gao, J.; Whitesides, G. M. Using Protein Charge Ladders to Estimate the Effective Charges and Molecular Weights of Proteins in Solution. *Anal. Chem.* **1997**, *69*, 575–580.
(29) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1802.
(30) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(31) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. Insights into Current Limitations of Density Functional theory. *Science* **2008**, *321*, 792–794.
(32) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural-Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.
(33) Guerra, C. F.; Handgraaf, J. W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, And VDD Methods for Charge Analysis. *J. Comput. Chem.* **2004**, *25*, 210.
(34) Xie, W. S.; Song, L. C.; Truhlar, D. G.; Gao, J. L. The Variational Explicit Polarization Potential and Analytical First Derivative of Energy: Towards a Next Generation Force Field. *J. Chem. Phys.* **2008**, *128*, 234108.
(35) Anisimov, V. M.; Bugaenko, V. L.; Cavasotto, C. N. Quantum Mechanical Dynamics of Charge Transfer in Ubiquitin in Aqueous Solution. *ChemPhysChem* **2009**, *10*, 3194–3196.