pK_a Calculations with QM/MM Free Energy Perturbations

Guohui Li and Qiang Cui*

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Ave, Madison, Wisconsin 53706

Received: June 8, 2003; In Final Form: August 23, 2003

A new approach for predicting the pK_a value for a specific residue in complex environments has been proposed. It is based on a combination of hybrid quantum mechanical/molecular mechanical (QM/MM) potential and the free energy perturbation (FEP) technique. With a specific thermodynamic cycle, the QM/MM-FEP protocol can be carried out for pK_a predictions taking advantage of the dual-topology−single-coordinate scheme proposed earlier for performing FEP calculations with QM/MM potentials. The new method has been tested for ethanethiol (CH_3CH_2SH) in solution. It was shown that, although the dominant contribution is from electrostatic interactions between the solute and solvent, many other factors have to be carefully dealt with to obtain reliable pK_a values. The contribution from van der Waals interactions associated with the dummy atom was found to be insignificant for the present case, because the solvent structure around the solute is essentially determined by the heavy atoms. With more systematic benchmark calculations, the QM/MM-FEP approach will become a useful tool for pK_a predictions in systems such as metalloenzymes, where nontrivial electronic structural change and/or atomic structural rearrangements are expected as the protonation state of a specific residue varies.

I. Introduction

The protonation state of titratable groups is critical to the functions of biological systems.[1] Changes in the protonation state due to pH or other perturbations in the surroundings (e.g., redox reaction of a nearby group) are important for the regulation of biomolecular processes. Therefore, it is of great interest to be able to predict the protonation states of specific residues and analyze how are they modulated by the environment. Our long-term goal is to understand the molecular mechanisms of molecular pumps involved in bioenergy transductions (e.g., cytochrome *c* oxidase[2−4]), where the coupling between redox reactions at the metal centers and changes in the protonation states of various residues are the key elements. Another good example where the pK_a of a specific group is essential to the chemistry is peptide synthesis in ribosome, where there remains controversy about the precise pK_a of Ade 2451 and whether the catalysis has a general acid−base mechanism or is due mainly to electrostatic effects.[5]

Not surprisingly, there has been a great deal of activities along this line in both the quantum chemistry and molecular simulation fields, albeit with somewhat different emphases and levels of treatment. In the quantum chemistry arena, the major motivation has been to predict highly accurate pK_a for a specific group. In the gas phase, this amounts to the prediction of highly accurate proton affinity, which can be achieved with highly correlated ab initio methods for small molecules[6,7] or the hybrid ONIOM approach for a small functional group in a large molecule.[8,9] In the condensed phase (solution or solvated macromolecule), solvation energies of the various species in the commonly used thermodynamic cycle (Scheme 1a) have to be taken into account, which has been done with continuum dielectric models or a combination of explicit solvent and continuum models.[10−14] It is not trivial to compute accurate solvation energies for charged species (including proton) with these approaches,[15,16] and therefore, predicting pK_a in solution remains a challenge for complex molecules.[17] In the field of molecular simulation, which mainly aims at predicting pK_a's for many ionizable groups in a solvated macromolecule (such as titratable amino acids in a protein), approximations have to be made based on physical considerations. The common practice is to assume that a change in the electrostatic interactions is the most important contribution as the protonation states vary; thus, electrostatic calculations such as those based on solving the Poisson−Boltzmann (PB) equation[18,19] are often used.[20−25] Because the bond dissociation energies are not taken into account in most PB based approaches, one constructs a different thermodynamic cycle (Scheme 1b) such that it is essentially the pK_a shifts in the macromolecule relative to isolated residues in solution that are estimated. Because the pK_a's for isolated residues in solution are known by experiments, the absolute pK_a's of these residues in the macromolecule can then be determined. The key elements[25] to the success of PB based calculations include the choice of the dielectric constant and atomic radii for the macromolecule,[23] as well as a sufficient average over multiple conformations.[26−28]

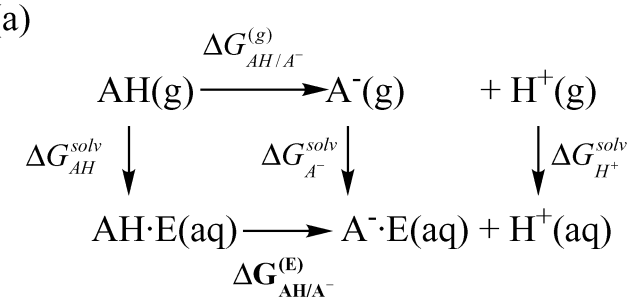
In the present study, we attempt to bridge the two fields with a combination of hybrid quantum mechanical/molecular mechanical (QM/MM) potentials[29−31] and free energy perturbation (FEP)[32−35] techniques. This is a natural extension of our previous work on using a similar approach to compute redox potentials in the macromolecular environment.[36,37] As microscopic models, FEP based approaches[33−35] have the advantage that changes in the interactions between the environment and the ionizable group

* To whom correspondence should be addressed.

10.1021/jp0356158 CCC: $25.00
Published on Web 11/25/2003

SCHEME 1

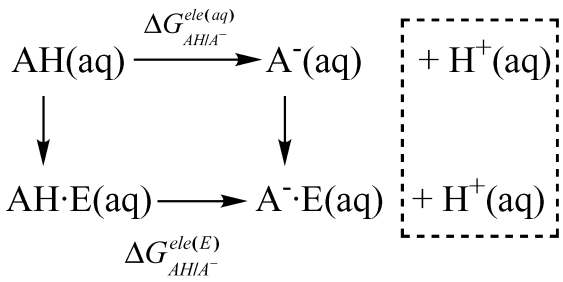
(a)



$$\Delta \mathbf{G}^{(\mathbf{E})}_{\mathbf{AH/A^-}} = \Delta G^{(g)}_{AH/A^-} + \Delta G^{solv}_{A^-} - \Delta G^{solv}_{AH} + \Delta G^{solv}_{H^+}$$

$$pK_a^{(E)} = \Delta \mathbf{G}^{(\mathbf{E})}_{\mathbf{AH/A^-}} \big/ (2.303RT)$$

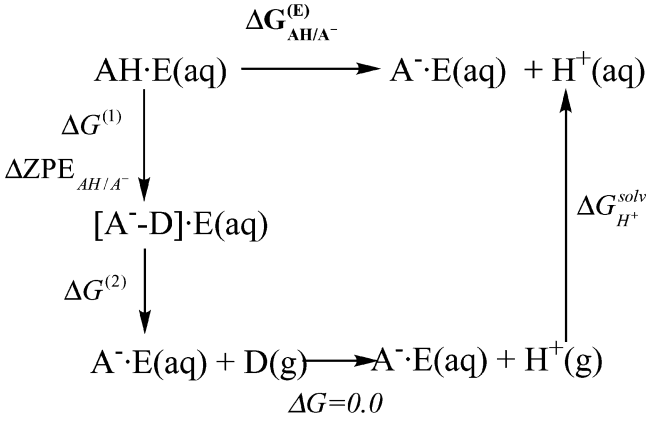
(b)



$$pK_a^{(E)} = pK_a^{(aq)} + \left[\Delta G^{ele(E)}_{AH/A^-} - \Delta G^{ele(aq)}_{AH/A^-}\right] \big/ (2.303RT)$$

in different protonation states are taken into account in a rigorous manner. For example, in many problems that we are primarily interested in (e.g., cytochrome *c* oxidase), where the ionizable group is attached to a metal ion, the local environment is expected to undergo significant rearrangements as the protonation state of the ionizable group varies, which have been neglected in most quantum chemical[10−12] and PB calculations.[20−23] In these cases, the importance of going beyond the single-structure framework inherent in the PB based methods has been emphasized by Warshel and co-workers,[28] especially in the context of redox potential calculations in protein environments.[38,39] From the computational point of view, the FEP approaches are robust because interactions *within* the environment, which typically are much larger in magnitude compared to the free energy difference of interest, do not contribute explicitly to the result and therefore numerical convergence is not a serious issue (not to imply that all studies in the literature pay equal attentions to convergence). In the context of pK_a calculations, FEP has been used by Warshel et al.,[17,40−42] Jorgensen et al.,[43] and Merz[44] with classical force fields and appropriate thermodynamic cycles; rather reliable results (e.g., ∼±3 pK_a values in ref 44) can be obtained for pK_a shifts relative to model systems in solution. The advantage of the present QM/MM-FEP approach is that it efficiently accounts for the differential response of the QM electronic structure to the environment when the QM region is in different chemical states (i.e., protonated and ionized states in the present study). Therefore, the method is potentially more robust than those based on classical point charge models,[17,40−42,44] especially for systems with complex electronic structures such as metalloenzymes and bioluminescent proteins.

SCHEME 2



$$\Delta \mathbf{G}^{(\mathbf{E})}_{\mathbf{AH/A^-}} = \Delta G^{(1)} + \Delta G^{(2)} + \Delta G^{solv}_{H^+} + \Delta \mathrm{ZPE}_{AH/A^-}$$

$$pK_a^{(E)} = \Delta \mathbf{G}^{(\mathbf{E})}_{\mathbf{AH/A^-}} \big/ (2.303RT)$$

In the following, we first describe the QM/MM-FEP approach and the specific thermodynamic cycle used for pK_a calculations. In section III, the method is tested with a small molecule, CH_3-CH_2SH, in solution. This is a model for the side chain of Cys, which is a typical ligand for metal ions such as Zn and Cu in several enzymes of interest; thus, understanding the pK_a of Cys is crucial in the mechanistic study of these systems. Finally, we summarize in section IV.

II. Theory and Computational Methods

To calculate the pK_a for a specific residue in a macromolecule in the condensed phase, a typical thermodynamic cycle that has been adopted in previous mixed quantum chemical/continuum calculations is shown in Scheme 1a.[10−12,40,41] Accordingly, one needs to compute the proton affinity of the residue in the gas phase (which includes contributions such as zero-point effects) and the differential "solvation free energy" for transferring the protonated and depronated species into the solution or macromolecular environment. The latter solvation free energies are often calculated with a continuum dielectric model[15,16] either in an approximate quantum mechanical framework (e.g., QM/Effective Fragment Potential)[10] or with a classical point charge parametrization (e.g., the residue is replaced by fitted point charges from gas-phase QM calculations).[11]

In the QM/MM-FEP approach, we use an alternative thermodynamic path as illustrated in Scheme 2, which differs from the popular ones (Scheme 1) in several aspects. Our thermodynamic path involves transforming the residue from the protonated form into the deprotonated form with a free energy perturbation (FEP) protocol. As discussed extensively in the past,[17,32−35,42] FEP has the advantage that one computes directly the *difference* in the "solute"-environment interaction when the "solute" is in two chemical states, which involves the statistical

average of small numbers. In the common PB-based methods, by contrast, large numbers such as the *total* electrostatic interactions are used to compute differential solvation free energies (Scheme 1). Therefore, a FEP based approach is more robust from the numerical stability point of view.

In a previous article,[37] we pointed out a number of complications associated with using the hybrid QM/MM potential in FEP calculations, which are mainly due to the fact that QM potentials are not separable and it is often difficult to define a fractional QM system. To circumvent these complications, we proposed a straightforward dual-topology−single-coordinate (DTSC) scheme, where the two chemical QM states adopt the same Cartesian coordinates during the perturbation process. We note that by "single-coordinate" we mean that two electronic structure calculations for the two chemical states (dual-topology) are carried out at each geometry (single-coordinate) during the FEP simulation. The geometry of the perturbed system is fully flexible during the simulation (except for bonds involving hydrogen, which are usually constrained using SHAKE[45]); as λ is switched from 0 to 1, the "single-coordinate" naturally switches from that which favors the initial state to that which favors the final state. Therefore, no error arises in the DTSC approach even if the two chemical states do not have identical geometries, and no Jacobian terms are needed for artificially switching bonded parameters.[46,47] As a matter of fact, because the free energy change is path-independent, the DTSC scheme is *formally exact*;[37] it can be regarded as a variant of the mapping potential approach introduced by Warshel and co-workers[48] in the context of FEP calculations for diabatic states in electron transfers. In practical simulations, small errors might arise if the bonds involving *hydrogens* have very different bond distances in the two end-states, because they are usually constrained to be one set of values during the MD simulations; the magnitude of such effects, however, is expected to be very small in most realistic applications. Taking the test pair of conjugate acid/base of $CH_3CH_2SH/CH_3CH_2S^-$ as an example, the optimal C−H distances differ by less than 0.01 Å between the protonated and deprotonated forms in the gas phase at the B3LYP/6-31+G(d,p) level; that is, constraining the C−H distances to those in the optimized protonated form raises the energy of $CH_3CH_2S^-$ by less than 0.08 kcal/mol. From a numerical point of view, the DTSC scheme has the advantage that the free energy derivative is very easy to compute and no new matrix elements have to be calculated associated with the QM/MM coupling. Moreover, the DTSC scheme is free of end-point problems.[49,50]

In its original form, the DTSC scheme clearly has to involve two chemical states with similar structures (i.e., the number of atoms and connectivities; the geometries do *not* have to be the same, as discussed in the above paragraph), which is not a serious limitation because otherwise either a QM/MM potential is not called for (e.g., when the two states are entirely different molecules) or other calculations (e.g., potential of mean force[51,52] for significant chemical rearrangements) are more useful. In pK_a calculations, to make it possible to adopt the DTSC scheme, we perform the "alchemy mutations" following a specific thermodynamic path (Scheme 2). First, we mutate the residue of interest from the protonated form (AH) into the deprotonated form with a dummy atom attached, which replaces the ionized proton (A^-−D); both AH and A^- will be treated with QM, and the environment (macromolecule plus solvent) will be treated with MM. The dummy atom (D) has the same mass and van der Waals parameters as the acidic proton in AH; it interacts with A^- through classical bonded terms (bond, angle, dihedral) and interacts with all other atoms with van der Waals (but no electrostatic) interactions. The introduction of the dummy atom is to conserve the number of degrees of freedom, and the adoption of the same mass and van der Waals parameters as the proton in AH makes the DTSC scheme possible. The classical bonded terms between A^- and the dummy atom are introduced to avoid sampling problems associated with a floating dummy atom[49,50] and will be removed later (see below). The dummy atom can be regarded as a proton with the nuclear charge set to zero; thus, the current DTSC scheme has the same end-states as a single-topology scheme in which one scales the nuclear charge associated with the H from 1.0 in AH to 0.0 in A^-−D (Gao, private communication). The advantage offered by DTSC is that the free energy derivative is very straightforward to compute, which involves only the QM/MM potential for the AH ($\lambda = 1.0$) and A^--D ($\lambda = 0.0$) states in the presence of the MM environment (E)

$$\frac{\partial G^{(1)}}{\partial \lambda} = \langle -U_{\mathrm{AH;E}} + U_{(\mathrm{A^{-}-D});\mathrm{E}} \rangle_\lambda = \langle -U_{\mathrm{AH;E}}^{\mathrm{QM/MM,ele}} + U_{\mathrm{A^{-}-D;E}}^{\mathrm{QM/MM,ele}} + U_{\mathrm{D}}^{\mathrm{Bonded}} \rangle_\lambda \quad (1)$$

where

$$U_{\mathrm{X;E}}^{\mathrm{QM/MM,ele}} = \langle \Psi_{\mathrm{X;E}} | \hat{\mathrm{H}}_{\mathrm{X}}^{\mathrm{QM}} + \hat{\mathrm{H}}_{\mathrm{X;E}}^{\mathrm{QM/MM,ele}} | \Psi_{X;E} \rangle \quad (2)$$

is the electrostatic component of the QM/MM potential for species X (AH or A^-−D) in the presence of the environment E.[30] As emphasized in earlier work,[36,37] van der Waals and bonded terms associated with the QM/MM interactions do *not* contribute to the free energy derivatives if the same set of parameters is used for the two chemical states; this is the reason that the same van der Waals parameters are used for the dummy atom in A^-−D as the acidic proton in AH. Such a convenient feature relies on the assumption that the same set of van der Waals and bonded parameters are appropriate for both chemical states, which is what often used in practical applications; although it is in principle possible to derive state-dependent parameters, it is unlikely that these parameters will be highly transferable. We note that bonded terms between the dummy atom and A^- ($U_{\mathrm{D}}^{\mathrm{Bonded}}$) also contribute to the free energy derivative, although in practice this term is very small (see section III).

In the second step along the thermodynamic path (Scheme 2), the dummy atom is transferred to the gas phase; that is, we need to compute the free energy change associated with switching off interactions between the dummy atom and the rest of the system (A^-;E). The dummy atom interacts with all other atoms through van der Waals interactions, and also with A^- through bonded terms. The former can be estimated with a second set of alchemy free energy perturbations in which van der Waals parameters associated with the dummy atom are gradually switched to zero. The bonded contributions can be estimated based on the local configurational integrals,[53] as discussed extensively in previous studies.[46,54] The free energy change associated with this step is given as

$$\Delta G^{(2)} = \Delta G_{\mathrm{D}}^{\mathrm{vdW}} + \Delta G_{\mathrm{D}}^{\mathrm{Bonded}} = -\int_0^1 \mathrm{d}\lambda \langle U_{\mathrm{D}}^{\mathrm{vdW}} \rangle_\lambda + \Delta G_{\mathrm{D}}^{\mathrm{Bonded}} \quad (3)$$

where $U_{\mathrm{D}}^{\mathrm{vdW}}$ is the van der Waals interaction between the dummy atom and all other atoms. Following the previous

discussions,[53,46,54] the bonded contribution is given by

$$\Delta G_{\mathrm{D}}^{\mathrm{Bonded}} = -k_{\mathrm{B}}T \ln \frac{V_0 \Lambda_D^{-1}}{r_{\mathrm{A}^-\mathrm{D}}^{2} \sin\theta_{\mathrm{A}^-\mathrm{D}} \sqrt{(2\pi k_{\mathrm{B}}T)^2/K_\theta K_\tau}} + \frac{5}{2}k_{\mathrm{B}}T \quad (4)$$

where $r_{A^-D}(\theta_{A^-D})$ is the distance (angle) that characterizes the connection between the dummy atom and A^-. Note that because the A^-−D bond is constrained using SHAKE,[45] only the bending and torsional force constant contribute and the kinetic piece contributes a square term rather than a cubic term in the denominator. The quantities V_0 and Λ_D are the molar volume and thermal wavelength associated with the dummy atom under standard state; Λ_D contributes in eq 4 because SHAKE was used for one degree of freedom (A^-−D bond).

Following the two alchemy free energy simulations, the system consists of a dummy atom in the gas phase and A^- "solvated" in the environment E. This differs from the appropriate final state by a solvated proton (Scheme 2); thus, the last missing component is essentially the solvation free energy of the proton ($\Delta G_{\mathrm{H}^+}^{\mathrm{solv}}$); the free energy associated with switching the dummy atom to a proton (or increasing the nuclear charge of the proton from 0 to +1) in the gas phase is zero.

Finally, we note that all of the calculations discussed above are classical simulations for the nuclear degrees of freedom (including the bonded contribution related to the dummy atom given in eq 4). Therefore, the difference in the quantum mechanical vibrational energies associated with AH and A^- states have to be included, and the simplest approximation is to use the zero-point energy difference between the two states in the gas phase ($\Delta \mathrm{ZPE}_{\mathrm{AH/A}^-}$).

Taking all of the contributions discussed above, the expression for pK_a in the present QM/MM-FEP approach is

$$\mathrm{p}K_a = \frac{\Delta G_{\mathrm{AH/A}^-}^{(\mathrm{E})}}{2.303RT} = \frac{1}{2.303RT}\left[\int_0^1 \frac{\partial G^{(1)}}{\partial \lambda} \mathrm{d}\lambda + \Delta G_{\mathrm{D}}^{\mathrm{vdW}} + \Delta G_{\mathrm{D}}^{\mathrm{Bonded}} + \Delta G_{\mathrm{H}^+}^{\mathrm{solv}} + \Delta \mathrm{ZPE}_{\mathrm{AH/A}^-}\right] \quad (5)$$

where $\partial G^{(1)}/\partial\lambda$ and $\Delta G_{\mathrm{D}}^{\mathrm{vdW/Bonded}}$ are given in eqs 1, 3, and 4, respectively; $\Delta G_{\mathrm{H}^+}^{\mathrm{solv}}$ has been estimated in previous work,[55−57] although there has been significant variations (see section III), and $\Delta \mathrm{ZPE}_{\mathrm{AH/A}-}$ is from gas-phase vibrational calculations for AH and A^-.

Before moving on to test calculations, we emphasize that the QM/MM-FEP approach is clearly not a very fast avenue toward pK_a predictions. It involves two sets of free energy perturbation calculations with hybrid QM/MM potentials, although test calculations in section III indicate that the second set makes a very small contribution and can often be neglected in practical applications. Therefore, it is *not* our intention to recommend using such an approach in *all* applications. In certain investigations where accurate pK_a for a specific group in a complex environment is of crucial mechanistic value and/or the change in the protonation state induces significant local structural rearrangements, on the other hand, the QM/MM-FEP approach is a very attractive and robust choice. In application to macromolecules, where it is crucial to ensure the convergence of FEP calculations, it is important to choose a QM level that is an appropriate compromise of accuracy and speed. Alternatively, one can choose a fast QM level or even classical potential functions as the reference potential in FEP calculations and then improve the results by doing perturbations using a higher-level QM method.[39,58]

TABLE 1: Gas Phase Proton Affinity (PA) and Zero-Point Vibrational Energies (ZPE) of CH_3CH_2SH at Different Levels of Calculations[a]

	SCC-DFTB[b]	B3LYP/I[c]	CCSD/I//B3LYP/I[c]	MP2/II//B3LYP/I[c]	CCSD/II//B3LYP/I[c]
total E	−8.11436	−478.07030	−477.28200	−477.34427	−477.38745
PA	355.5	358.7	366.2	364.8	368.1
total ZPE	45.9	46.7			
ΔZPE	6.0	6.3			

[a] The total potential energies of CH_3CH_2SH (total E) are given in hartree, all other quantities (proton affinity of CH_3CH_2SH (PA), total zero-point energy of CH_3CH_2SH (Total ZPE), and the ZPE difference between CH_3CH_2SH and $CH_3CH_2S^-$ (ΔZPE)) are given in kcal/mol. [b] A previously established value of 141.8 kcal/mol was used as the reference for proton (see ref 76 for more details). [c] Basis set I is 6-311+G(d,p), and basis set II is cc-pVTZ.

III. Test Calculations on CH_3CH_2SH and Analysis

In this section, the newly proposed QM/MM-FEP approach is tested with a relatively simple system: CH_3CH_2SH in aqueous solution. The molecule is a realistic model for Cys side chain; it is chosen here because the protonation state of Cys is important in several metalloenzymes of interest.[59,60] In addition to testing the reliability of the QM/MM-FEP approach for predicting pK_a values, it is also of interest to explore the relative weights of various terms in eq 5 and determine if any contribution can be approximated or even safely ignored.

A CH_3CH_2SH molecule is solvated in a 32.0 Å cubic box of water molecules, where the solution is treated with an approximate density functional theory (SCC-DFTB),[61] and the water molecules are described with the standard TIP3P model,[62] slightly modified in CHARMM.[63] The SCC-DFTB method is chosen because it is computationally efficient and is rather accurate.[64−67] As seen in Table 1, it predicts the gas-phase proton dissociation energy of CH_3CH_2SH very close to density functional theory (B3LYP) calculations, although both SCC-DFTB and B3LYP gave too low proton affinities compared to CCSD/cc-pVTZ.[68] For the interaction between SCC-DFTB and TIP3P, both electrostatic and van der Waals interactions are considered.[67] The van der Waals parameters for the SCC-DFTB atoms are taken from the CHARMM 22 force field[69] for the same molecule (see the Supporting Information). The free energy perturbation calculations have been performed with a modified version of CHARMM,[70] which allows the DTSC calculations and periodic boundary condition MD simulations using the SCC-DFTB/MM potential. For bonded interactions between the dummy atom and $CH_3CH_2S^-$, the angle bending parameters are taken from CHARMM 22 force field, and the dihedral angle was described by a very soft harmonic term (see footnote of Table 2). The bond stretching term for the dummy atom is irrelevant because all bonds involving hydrogen atoms and the dummy atom are constrained during the MD-FEP simulation with SHAKE.[45] In the MD-FEP simulations, the integration step size was 1 fs, and the nonbonded cutoff scheme is 12−10Å, although calculations with 10−8 Å cutoff scheme were also used to test the reliability of the results. In the cutoff scheme, the two numbers indicate the range in which nonbonded interactions were modified based on a force-shift algorithm.[71]

All contributions to pK_a (eq 5) in the QM/MM-FEP approach are listed in Table 2. The short conclusion is that it is challenging to predict accurate pK_a, and one has to be careful with the

TABLE 2: Various Contributions to the pK_a of CH_3CH_2SH in Solution in the SCC-DFTB/TIP3P−FEP Approach[a]

contributions	values
$\Delta G^{(1)}$ (eq 1)[b]	285.0/288.1
$\Delta G^{(1)}$ Born correction[b]	−13.8/−16.6
ΔG_D^{vdW} (eq 3)	0.05
ΔG_D^{Bonded} (eq 4)[c]	−3.7
$\Delta ZPE_{AH/A^-}$[d]	−6.0
$\Delta G_{H^+}^{solv}$[e]	−251.0 (−265.0) [−262.4]
$\Delta G_{AH/A^-}^{(E)}$ (eq 5)[f]	10.6 (−3.5) [−0.9]
pK_a[f]	7.7 (−2.5) [−0.7]
corrected pK_a[g]	16.8 (6.6) [8.5]
experimental pK_a[h]	10.6

[a] All free energies and zero-point energy corrections (ΔZPE) are given in units of kcal/mol. [b] The numbers before and after the slash were calculated with the cutoff scheme of 12−10 and 10−8 Å, respectively; in the cutoff scheme, the two numbers indicate the range in which the nonbonded interactions were modified based a force-shift algorithm.[71] For the Born correction, the formula $q^2/2R((\epsilon - 1)/(\epsilon))$ was used,[73] where $q = -1$, $\epsilon = 80$, and R corresponds to the larger value in the cutoff scheme used in the free energy perturbation calculations (i.e., 12 and 10 Å in the two cutoff schemes, respectively). [c] The following numerical values are used:[53] $V_0 = 136.22$ T/P (Å^3), $\Lambda_D = 17.457(mT)^{-1/2}$ (Å), $T = 300.0$ K, $P = 1.0$ atm, $m = 1.008$ g/mol, $K_\theta = 38.8$ kcal/(mol rad^2), $K_\tau = 0.025$ kcal/(mol rad^2). The relevant dummy−$CH_3CH_2S^-$ geometrical parameters in eq 3 are $r_{A^--D} = 1.325$ Å, $\theta_{A^--D} = 95°$. [d] Gas-phase value in Table 1 was used. [e] The extreme values are taken from ref 55, and the value of −262.4 kcal/mol (in brackets) was taken from high-level ab initio calculations in ref 57. [f] The three values were computed using the three corresponding proton solvation free energies ($\Delta G_{H^+}^{solv}$) cited in footnote e. [g] A correction was used for SCC-DFTB based on the discrepancy found for CH_3CH_2SH compared to CCSD/cc-pVTZ results in the gas phase (see Table 1). The three values were computed using the three corresponding proton solvation free energies ($\Delta G_{H^+}^{solv}$) cited in footnote e. [h] From ref 72.

various contributions; one pK_a unit corresponds to 1.37 kcal/mol difference in free energy. Compared to the experimental value of 10.6,[72] the closet estimate from the current work (*without* optimization of any specific calculation parameters) is 8.5. Part of the difficulty is due to the uncertainty in the proton solvation free energy, and a practical approach will be establishing the appropriate value to use with QM/MM-FEP through more systematic calculations for a large number of chemical species with well-established pK_a values, which are being carried out in our group. In the following, we analyze the various contributions in more detail and comment on their importance to the reliability of the predicted pK_a value.

Not surprisingly, the dominant contribution in the QM/MM-FEP approach using the thermodynamic cycle in Scheme 2 is from the first step where AH is mutated into A^-−D in solution. During this process, the A−H bond is broken, which costs energy; on the other hand, A^- is better solvated than AH; thus, the free energy change is substantially reduced compared to the proton affinity of A^- in the gas phase (Table 1). As shown in Figure 1a, the free energy derivatives are all well-behaved, including those *at* the end-points ($\lambda = 0$ and 1), which is a nice feature of the DTSC scheme.[36,37] Most windows approach convergence after 30 ps of production MD calculations, although certainly windows take longer to converge; the $\lambda = 1$ simulation has fairly significant variations even after 200 ps. As to the λ-dependence of the free energy derivatives, the striking result is that the relationship is highly linear ($R^2 = 0.999$; Figure 1b). This is, once again, a unique feature of the DTSC scheme, in which the free energy derivatives are dominated by electrostatic interactions, which follow the linear response very faithfully. We note that, because of the electrostatic nature of $\Delta G^{(1)}$, it is crucial to take the long-range character of electrostatic interactions into account. In the current work, where a cutoff scheme is used for all electrostatic interactions (including that between QM and MM atoms), we used the Born model[73] to make a correction. As shown in Table 2, the Born correction is substantial (∼−13.8 kcal/mol) and makes a significant contribution to the accuracy of the calculated pK_a value. To test the robustness of the electrostatic treatment, $\Delta G^{(1)}$ was also calculated with a shorter cutoff of 10−8 Å; without the Born correction, the $\Delta G^{(1)}$ was ∼3.2 kcal/mol higher than the value calculated with the 12−10 Å cutoff (see Table 2). However, it is encouraging to see that such an increase is almost compensated by the Born correction. When the Born correction is included, $\Delta G^{(1)}$ is ∼272 kcal/mol with both cutoff schemes (see Table 2). Therefore, we can conclude that combining the cutoff with a Born correction is an effective scheme to describe electrostatic interactions in pK_a calculations, although the Ewald summation is still a desirable option for the future.

For the second alchemy free energy simulation, in which the van der Waals parameters associated with the dummy atom are switched to zero, the free energy derivatives are surprisingly small in magnitude (absolute value <1 kcal/mol; Figure 1c), and the final integrated contribution is only 0.05 kcal/mol (Table 2). This is somewhat unexpected because previous simulations[49,74] and scaled-particle theory[75] argued that free energy derivatives are expected to diverge at the end point ($\lambda = 1$), where the environment particles can approach the scaled particle very closely as the van der Waals interactions between them are scaled by $1 - \lambda$. The present results can be understood by examining the radial distribution functions of the solvent around the dummy atom. As shown in Figure 2, the solvent distribution actually changed rather little as the van der Waals parameters associated with the dummy atom are switched off; the peaks in the distribution functions shift inward by only 0.3−0.4 Å and still are far from the center of the dummy atom. This is because, in contrast to previous analysis,[75] the dummy atom is not by itself but is bonded to A^- in the present simulations, and the van der Waals interaction between A^- (or S in $CH_3CH_2S^-$) and the solvent is sufficient to prevent the latter from collapsing onto the dummy atom even if the dummy atom has scaled van der Waals parameters. This resembles the situation in conventional water models (e.g., the standard TIP3P[62]), where no van der Waals parameters are needed for the hydrogen atoms to prevent water molecules from collapsing onto each other because the interactions between oxygen atoms are sufficient to determine the bulk water structure. Therefore, the current test calculation suggests that the van der Waals contribution due to the dummy atom is fairly insignificant, and therefore the second alchemy free energy simulation can often be safely skipped. Benchmark calculations are required for more heterogeneous systems such as proteins.

As shown in Table 2, the other two contributions, the bonded interaction between the dummy atom and A^- (ΔG_D^{Bonded}) as well as the ZPE difference between AH and A^- ($\Delta ZPE_{AH/A^-}$) are also important. They contribute −3.7 and −6.0 kcal/mol, respectively, to the total free energy difference, $\Delta G_{AH/A^-}^{(E)}$. Although the solvation free energy for the proton is a constant, it is large in magnitude and actually fairly difficult to probe either experimentally or computationally.[55−57] There are significant variations ranging from −251 to −265 kcal/mol, and a recent high-level ab initio calculation study[55−57] recommended the value of −262.4 kcal/mol; using the three values gives significantly different pK_a results (Table 2). In practice, the best approach is to perform QM/MM-FEP calculations for a series

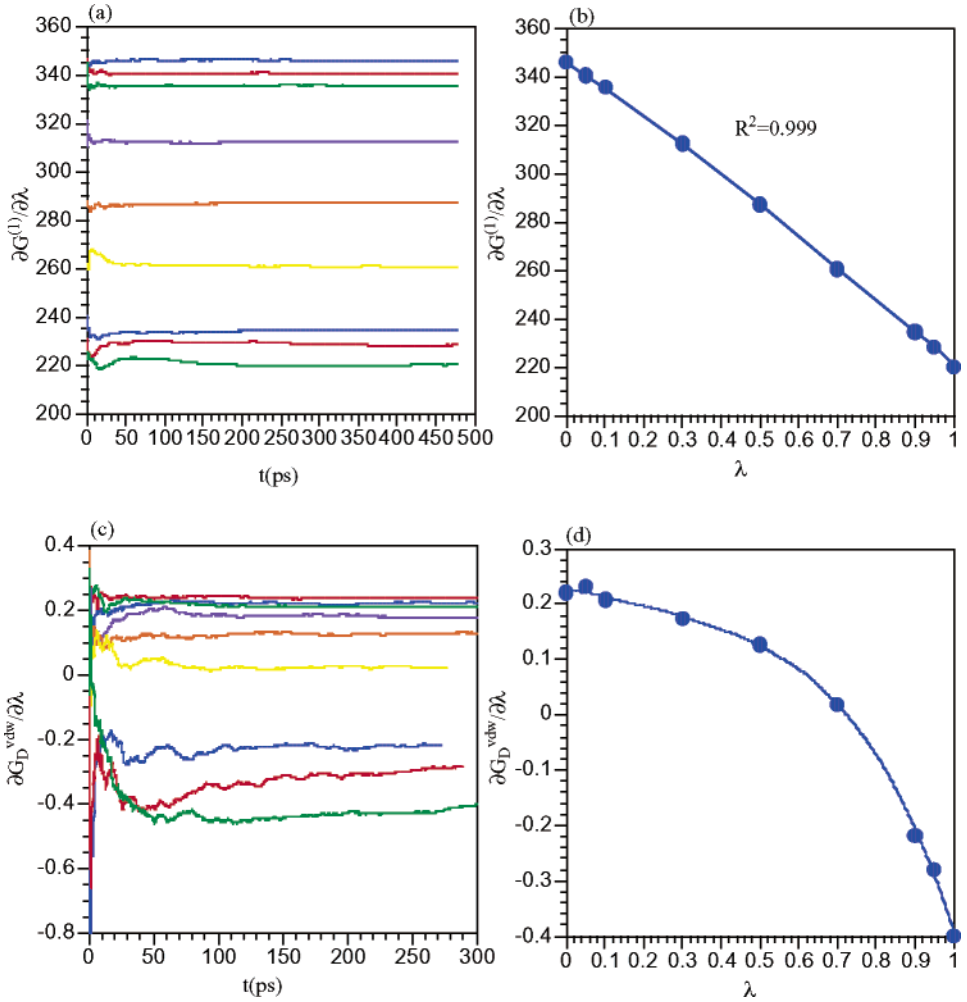


Figure 1. Free energy derivatives associated with the two sets of QM/MM free energy perturbation calculations for the pK_a calculation of CH_3CH_2SH in solution. As shown in Scheme 2, the first set (a,b) involves mutating CH_3CH_2SH to $CH_3CH_2S^-$−D, where D is a dummy atom that replaces the acidic proton; both CH_3CH_2SH and $CH_3CH_2S^-$ were treated with the SCC-DFTB approach in the DTSC scheme (see text). In the second set (c,d), the van der Waals parameters on the dummy atom in $CH_3CH_2S^-$−D are gradually switched off; the $CH_3CH_2S^-$ was treated with SCC-DFTB. Note the dramatic difference in scales.

of compounds that have well-established pK_as' and then decide what value should be used for $\Delta G_{H^+}^{solv}$ for a specific QM/MM setup; i.e., we propose to use $\Delta G_{H^+}^{solv}$ as an adjustable parameter to partially account for the deficiency of QM/MM models. For CH_3CH_2SH, it first appears that the value of −251.0 kcal/mol is more appropriate, which gives a pK_a value of 7.7 for CH_3CH_2SH; the experimental estimate is 10.6.[72] However, gas-phase calculations indicate that, although SCC-DFTB agrees well with B3LYP calculations, both methods give too low proton affinities compared to CCSD/cc-pVTZ,[68] which is often regarded as systematically more reliable. Taking the difference between SCC-DFTB and CCSD/cc-pVTZ for the gas-phase model, we see that the value of −262.4 kcal/mol for $\Delta G_{H^+}^{solv}$ gave the pK_a (8.5) that is the closest to the experimental measurement.

In short, the test calculations clearly demonstrated that the dominant contribution concerns the difference in the electrostatic interactions between AH and A^- with the environment; thus, the treatment of long-range electrostatic interactions is crucial to the success of the QM/MM-FEP approach. The contribution from the van der Waals interactions associated with the dummy atom was found to be insignificant, and the bonded interactions with A^- are more important. To obtain accurate pK_a values, it is also essential to include the ZPE differences and to establish the appropriate value of $\Delta G_{H^+}^{solv}$ to use in eq 5. Although there are still uncertainties in the best set of parameters to use in the QM/MM-FEP approach, it is encouraging to see that, with a recent estimate of $\Delta G_{H^+}^{solv}$,[57] we are able to obtain a computed pK_a for CH_3CH_2SH in solution that is fairly close to the experimental value. Clearly, more systematic benchmark calculations are required to establish the best QM/MM-FEP protocol for pK_a predictions. We note that $\Delta G_{H^+}^{solv}$ do not contribute if only the pK_a shift between different residues or the same residue in different environments (e.g., solution vs protein) is of interest.

IV. Conclusions

We presented a new approach for computing pK_a for a specific residue in complex environments based on a combination of hybrid QM/MM potential and the free energy perturbation (FEP) technique.[17] Compared to previous methods based on the combination of continuum electrostatics and quantum mechanical[10−12] or classical mechanical potentials,[20−23] the QM/MM-FEP approach is more robust because it can handle nontrivial variations in both the electronic structure and nuclear geometry when the protonation state of the specific residue changes.[28,39]

The current approach is made possible by our recent development of a dual-topology−single-coordinate (DTSC) scheme for performing free energy perturbation calculations with hybrid

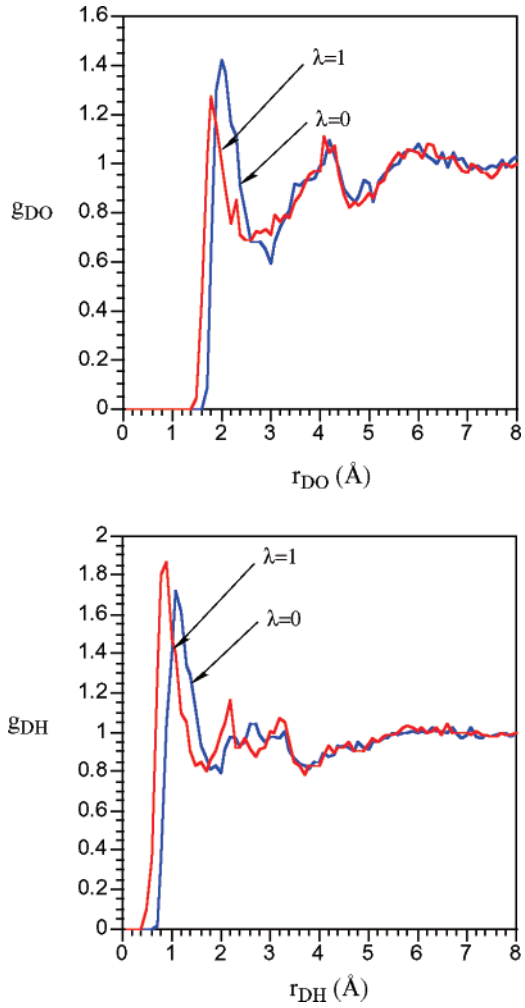


Figure 2. Radial distribution function for solvent oxygen and hydrogen atoms around the dummy atom at $\lambda = 0$ and 1 in the second set of alchemy free energy simulations in which the van der Waals parameters on the dummy atom are switched off ($\lambda = 1$). Note that the change is very small with different λ values, indicating that the van der Waals interactions between $CH_3CH_2S^-$ and solvent determine the solvent distribution, and the dummy atom makes little contribution.

QM/MM potentials.[36,37] With an appropriately constructed thermodynamic pathway (Scheme 2), the DTSC scheme has been successfully used for pK_a calculations. Formally, pK_a predictions with the QM/MM-FEP approach involves two sets of free energy perturbation calculations, although test calculations for CH_3CH_2SH in solution clearly demonstrated that the second set of FEP, which is related to the van der Waals contribution from the dummy atom that replaces the acidic proton, can be safely skipped due to its insignificant contribution; this conclusion needs to be verified for macromolecular systems with similar benchmark calculations. Because the pK_a is dominated by the difference in the electrostatic interactions between AH and A^- with the environment, test calculations also underlined the importance of an appropriate treatment of long-range electrostatics. The linear dependence of the dominant contribution found in the test calculations is consistent with the success of continuum dielectric based approaches, although the QM/MM-FEP approach is more robust for heterogeneous systems such as proteins and cases where significant structural changes accompany the change in the protonation state.

Finally, we emphasize once more that the QM/MM-FEP approach is not designed to predict pK_a for general protein systems because it is more expensive than the traditional PB based methods and the accuracy also involves the choice of a number of parameters (e.g., the accuracy of the QM approach and QM/MM interactions as well as the value of $\Delta G^{solv}_{H^+}$); it remains challenging to deal with multiple titration sites[20] with the QM/MM-FEP approach. Nevertheless, for certain systems such as proton pumps where it is essential to understand the pK_a's of specific groups and how they are coupled to other chemical processes (e.g., redox reactions), the QM/MM-FEP framework is very promising. The most important aspect for successful QM/MM-FEP applications to macromolecular systems is the balance of calculation speed (i.e., convergence of FEP) and accuracy, and the most productive way is to use an efficient QM level in the FEP calculations (e.g., with SCC-DFTB/MM, it is straightforward to carry out nanosecond FEP calculations[36,37]) and then improve the results by perturbation calculations[58] with carefully benchmarked high-level QM methods (e.g., MP2 or B3LYP). Such applications are in progress.

Acknowledgment. The Q.C. group is partially supported by the starting-up fund from the Department of Chemistry and College of Letters and Science at University of Wisconsin, Madison, a PRF-G grant administered by the American Chemical Society, and a Research Innovation Award from the Research Corporation. We thank Dr. D. Khoroshun for interesting discussions.

Supporting Information Available: The CHARMM RTF files for $CH_3CH_2SH/CH_3CH_2S^-$ are included. The van der Waals parameters are also listed. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular biology of the Cell*, 3rd ed.; Garland Publishing: New York, 1994.
(2) Michel, H. *Biochem.* **1999**, *38*, 15129−15140.
(3) Zaslavsky, D.; Gennis, R. B. *Biochim. Biophys. Acta* **2000**, *1485*, 164−179.
(4) Wikstrom, M. *Curr. Opin. Struct. Biol.* **1998**, *8*, 480−488.
(5) Jenni, S.; Ban, N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 212−219.
(6) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.
(7) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.
(8) Forese, R. D. J.; Morokuma, K. *J. Phys. Chem. A* **1999**, *103*, 4580−4586.
(9) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654.
(10) Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486-3494.
(11) Konecny, R.; Li, J.; Fisher, C. L.; Dillet, V.; Bashford, D.; Noodleman, D. A. *Inorg. Chem.* **1999**, *38*, 940.
(12) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610−5620.
(13) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413−7422.
(14) Klicic, J. J.; Friesner, R. A.; Liu, S. Y.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327−1335.
(15) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(16) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(17) Warshel, A. *Computer modeling of chemical reactions in enzymes and solutions*; Wiley & Sons: New York, 1997.
(18) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1975.
(19) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(20) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(21) Yang, A. S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 252.
(22) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415.
(23) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400.
(24) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.

(25) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819−7833.
(26) van Vlijmen, H. W. T.; Schaefer, M.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 145−158.
(27) Langen, R.; Brayer, G. D.; Berghuis, A. M.; McLendon, G.; Sherman, F.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 589.
(28) Sham, Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.
(29) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitza, K. B., Boyd, D. B., Eds.; VCH: New York, 1996; Vol. 7, p 119.
(30) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(31) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.
(32) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184−189.
(33) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.
(34) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.
(35) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.
(36) Formaneck, M. S.; Li, G.; Zhang, X.; Cui, Q. *J. Theor. Comput. Chem.* **2002**, *1*, 53−67.
(37) Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, In press.
(38) Warshel, A.; Papazyan, A.; Muegge, I. *J. Biol. Inorg. Chem.* **1997**, *2*, 143−152.
(39) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025−5039.
(40) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640.
(41) Warshel, A.; Weiss, R. M. *Biochemistry* **1981**, *20*, 3167.
(42) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368.
(43) Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.* **1989**, *111*, 4190−4197.
(44) Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 3572−3575.
(45) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.
(46) Boresch, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 5145.
(47) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2702.
(48) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.
(49) Roux, B.; Nina, M.; Pomes, R.; Smith, J. *Biophys. J.* **1996**, *71*, 670.
(50) Simonson, T. In *Computational biochemistry and biophysics*; Becker, O. M., Mackerell, A. D. J., Roux, B., Watanabe, M., Eds.; Marcel Dekker: New York, 2001.
(51) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(52) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865−880.
(53) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652−1661.
(54) Boresch, S. *Mol. Simul.* **2002**, *28*, 13−37.
(55) Grabowski, P.; Riccardi, D.; Gomez, M. A.; Asthagiri, D.; Pratt, L. R. *J. Phys. Chem. A* **2002**, *106*, 9145−9148.
(56) Tissandier, M.; Cowen, K.; Feng, W.; Gundlach, E.; Cohen, M.; Earhart, A.; Coe, J. *J. Phys. Chem. A* **1998**, *102*, 7787−7794.
(57) Zhan, C.; Dixon, D. *J. Phys. Chem. A* **2002**, *105*, 11534−11540.
(58) Strajbl, M.; Hong, G. Y.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333−13343.
(59) Valentine, J. S.; Gralla, E. B. *Copper containing proteins*; Academic Press: New York, 2002; Vol. 60.
(60) Simonson, T.; Calimet, N. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 37−48.
(61) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seigert, G. *Phys. Rev.* **1998**, *B58*, 7260−7268.
(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(63) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(64) Bohr, H. G.; Jalkanen, K. J.; Elstner, M.; Frimand, K.; Suhai, S. *Chem. Phys.* **1999**, *246*, 13−36.
(65) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2000**, *256*, 15−27.
(66) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2001**, *263*, 203-219.
(67) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.
(68) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
(69) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; 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.; Kraplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(70) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(71) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.
(72) Streitweiser, J. A.; Heathcock, C. *Introduction to organic chemistry*; Macmillan: New York, 1981.
(73) Born, M. *Z. Phys.* **1920**, *1*, 45.
(74) van Gunsteren, W.; Beutler, T.; Fraternali, F.; King, P.; Mark, A.; Smith, P. In *Computer simulation of biomolecules systems*; van Gunsteren, W., Weiner, P., Wilkinson, A., Eds.; ESCOM Science: Leiden, 1993.
(75) Simonson, T. *Mol. Phys.* **1993**, *80*, 441.
(76) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565−581.