The Prediction of Protein p K a 's Using QM/MM: The p K a of Lysine 55 in Turkey Ovomucoid Third Domain

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · APRIL 2002

Impact Factor: 3.3 · DOI: 10.1021/jp013995w

The Prediction of Protein pK_a's Using QM/MM: The pK_a of Lysine 55 in Turkey Ovomucoid Third Domain

Hui Li,[†] Alexander W. Hains,[†,#] Joshua E. Everts,[†] Andrew D. Robertson,[‡] and Jan H. Jensen*[,†]

Department of Chemistry and Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242

Received: October 30, 2001; In Final Form: January 14, 2002

A computational methodology for pK_a predictions of small molecules based on an ab initio quantum mechanics (QM) description of the acid and a linearized Poisson−Boltzmann equation (LPBE) description of bulk solvation using the polarized continuum method is presented. This QM/LPBE method is capable of reproducing the pK_a's of several functional groups found in amino acid residues with a root-mean-square deviation from experiment of 0.6 pH units. The practical applicability of the QM/LPBE method is extended to proteins by using a QM description of the ionizable residue and an effective fragment potential (MM) description of the rest of the protein. This QM/MM/LPBE method is used to predict a pK_a of the Lys55 residue in turkey ovomucoid third domain of 11.0, which is in good agreement with the experimental value of 11.1.

I. Introduction

Much of the chemistry of biomolecules is intimately tied to their protonation states and, therefore, to the pK_a's of the ionizable functional groups in the molecules. For example, the pK_a's of amino acid residues determine the pH dependence of protein stability and the catalytic mechanism of many enzymes, while the protonation state of many enzyme substrates or inhibitors determines their binding constants, solubility, and uptake by cells. Thus, accurate predictions of pK_a's will significantly aid in the rational design of new proteins, biocatalysts, and drugs.

The most popular computational pK_a-prediction methods for small molecules are based on the Hammet−Taft (HT) approach[1] in which substituent effects on pK_a's are assumed to be additive and the necessary parameters are derived from experimental pK_a measurements. The HT approach shares the advantages and disadvantages of purely empirical models: very efficient and accurate predictions of pK_a's for systems similar to the ones included in the parametrization but decreased accuracy for other molecules. Furthermore, the HT approach does not explicitly treat conformational effects and is not generally applicable to large systems such as proteins.

Most pK_a-prediction methods for proteins[2] are based on a thermodynamic cycle (Figure 1) whereby the protein pK_a is related to the experimentally determined pK_a of a model compound (pK_{model}), shifted by the change in the "environment energy" ($\Delta\Delta G_{\mathrm{env}}$). The term ΔG_{env} contains contributions from desolvation and interactions of the ionizable group with the rest of the protein. The most popular methods evaluate the interaction energy as the interaction of the molecular mechanics (MM) charges in the model system interacting with the electrostatic potential of the protein, obtained by numerically solving the linearized Poisson−Boltzmann equation (LPBE). The dielectric constant of the bulk solvent region is taken to be 80 and a lower (typically 4 or 20) dielectric constant is used for the protein interior.[3,4] If there are several titratable sites (with similar pK_a's) their pH-dependent charges must be taken into account in a self-consistent fashion (requiring several LPBE solutions), leading to an apparent pK_a.[5] A pK_a calculated without this term (i.e., where all other titratable sites are in their neutral protonation state) is referred to as the intrinsic pK_a.

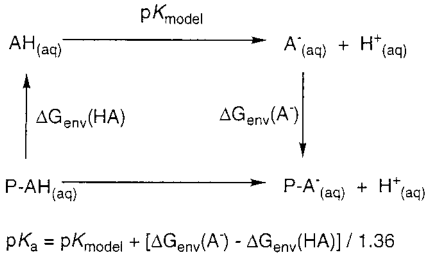

pK_{model}
$AH_{(aq)}$
$A^-_{(aq)}$ + $H^+_{(aq)}$
$\Delta G_{\mathrm{env}}(HA)$
$\Delta G_{\mathrm{env}}(A^-)$
P-$AH_{(aq)}$
P-$A^-_{(aq)}$ + $H^+_{(aq)}$
pK_a = pK_{model} + [$\Delta G_{\mathrm{env}}(A^-)$ - $\Delta G_{\mathrm{env}}(HA)$] / 1.36

Figure 1. Thermodynamic cycle relating the pK_a of a model compound (pK_{model}) to the pK_a of a protein residue via the environmental energies (ΔG_{env}) of the products and reactants. The value 1.36 corresponds to RT ln(10) at 298 K in kcal/mol.

The performance of the MM/LPBE approach for the prediction of protein pK_a's is generally good, with a root-mean-square deviation from experiment of typically <1 pH unit. However, larger errors are not uncommon, especially for residues with pK_a's significantly different than pK_{model}, not on the protein surface or both.[4]

For smaller systems the LPBE approach can be combined with a QM description of the molecule to yield absolute pK_a's, via another thermodynamic cycle (Figure 2). Lim, Bashford, and Karplus[6] were among the first to try this QM/LPBE approach, by using RHF/6-31G* proton affinities (without

* To whom correspondence should be addressed. E-mail: jan-jensen@uiowa.edu.

[†] Department of Chemistry.

[#] Current address: Department of Chemistry, University of Massachusetts, Amherst, MA 01003.

[‡] Department of Biochemistry.

10.1021/jp013995w CCC: $22.00
Published on Web 03/07/2002

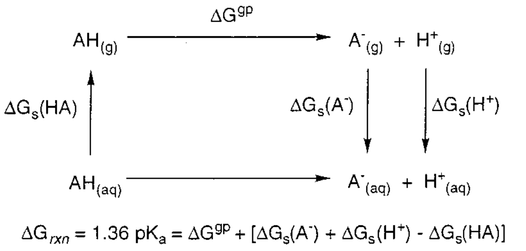


Figure 2. Thermodynamic cycle relating the pK_a to the gas-phase proton basicity (ΔG^{gp}) via the solvation energies (ΔG_s) of the products and reactants. The value 1.36 corresponds to RT ln(10) at 298 K in kcal/mol.

vibrational corrections) and a point-charge representation of the solute (Mulliken, CHARMM, or OPLS charges) for the LPBE solution. Using this approach, the calculated pK_a's were roughly twice the experimental values.

However, Chen et al.[7] subsequently demonstrated that the use of correlated methods (DFT) to calculate ΔG^{gp} combined with a self-consistent calculation of ΔG_s, in which the solute charges respond to the dielectric continuum, yields better results for several small compounds with key biomolecular functional groups[10] (Table 2). Further studies by Topol and co-workers[9] and Jang et al.[10] using very similar approaches have yielded pK_a's within 1 pH unit of experiment for a diverse set of molecules, including anti-HIV drug candidates and nucleotides.

Recently, Schüürmann et al.[11] used Tomasi's polarized continuum model[12] (PCM), in which the solute–continuum interaction is treated fully quantum mechanically, with modest results (possibly due to a methodological error). Subsequent studies by da Silva et al.[13] using H_2O/H_3O^+ rather than H^+ and PCM-optimized geometries gave much better results, as did similar studies by Shields[14] using the conductor model version of PCM and gas-phase geometries.

Thus, the intrinsic accuracy of the QM/LPBE approach is sufficient to compute protein pK_a's from first principles, provided that the computational cost for larger systems can be addressed (as well as conformational/dynamical effects and interactions between multiple titration sites). A QM/MM method offers an attractive solution because modeling the ionization equilibrium of a residue is unlikely to require a QM treatment of the entire system.

The concept of QM/MM was originally introduced by Warshel and Levitt[15] as an efficient way of including the effect of the protein environment (represented by a MM force field) on a QM model of enzyme catalysis. The application of electronic structure methods to biocatalysis through this approach was popularized by Bash et al.'s pioneering study of the reaction pathway of triosephosphate isomerase.[16] There are now many QM/MM methods available.[17] They differ mainly in their choice of QM (DFT/HF or AM1/PM3), MM force field (AMBER, CHARMM, GROMOS, or OPLS-AA), and their treatment of covalent QM/MM boundaries (link atom[18] or local-SCF[15,19]). Moreover, the treatment of solvation (explicit or implicit models) and free-energy effects (molecular dynamics or harmonic vibrational analysis) can also vary from one method to another.

Most biomolecular QM/MM applications have focused on the elucidation of enzyme mechanisms by computing the structure and relative energies of transition states and intermediate structures.[17] But Warshel and co-workers have also used the empirical valence bond-QM/MM method to compute relative pK_a values,[20] while Bash et al.[21] and Byun et al.[22] have also used AM1-QM/MM for small molecule pK_a predictions (where MM is an explicit solvation model). However, protein pK_a predictions using ab initio-QM/MM methods have not appeared in the literature.

Here, we describe the extension of the QM/LPBE approach for pK_a prediction to protein pK_a's by using a QM/MM description of the protein. The method is applied to the prediction of the pK_a of lysine 55 in turkey ovomucoid third domain (OMTKY3), a 56-residue protease inhibitor. This protein is unusually stable to pH, and the pK_a's of all of the ionizable residues have been measured *and assigned* using two-dimensional NMR spectroscopy by Robertson and co-workers.[23,24] The paper is organized as follows.

First, a QM/LPBE methodology, based on the PCM method, is presented. This method yields pK_a's within 0.9 pH units of experiment for small-molecule models of ionizable groups in proteins (acetic acid, methylamine, imidazole, phenol, and ethanethiol). Second, we briefly describe the effective fragment potential (EFP) method,[25] the QM/MM method on which our approach is based, including the recently developed interface with the PCM due to Bandyopadhyay, Gordon, Mennucci, and Tomasi,[26] and its extension to protein-sized system. Third, the application of the QM/MM/LPBE methodology to the calculation of the pK_a's of Lys55 in OMTKY3 is described. The minimization of the error incurred by using a single ionizable state is discussed in some detail. Fourth, we summarize our findings and describe future directions.

II. Computational Methodology

A. Small Molecule pK_a Predictions. On the basis of previous work,[6−11,13,14] we have developed the following methodology for small molecule pK_a predictions (cf. Figure 2). ΔG^{gp} is calculated at the MP2/6-31+G(2d,p)//RHF/6-31G(d) level of theory,[27] using frequencies scaled by 0.89 for the vibrational free energy correction (ΔG^{vib}; ΔG^{trans} and ΔG^{rot} are the translation and rotational free energies, respectively),

$$\Delta G^{gp} = \Delta E^{MP2//RHF} + \Delta G^{trans} + \Delta G^{rot} + \Delta G^{vib} + RT\ln(\tilde{R}T) \quad (1)$$

The last term in eq 1 changes the reference state from 1 atm to 1 M [$K_c = K_p(\tilde{R}T)$ for AH → $A^- + H^+$ reactions, where $\tilde{R} =$ 0.082 06 L atm/(mol K)].[14a]

The solvation energy is calculated by Tomasi's polarized continuum model (PCM[12]) using the united atom for Hartree–Fock (UAHF) radii proposed by Barone, Cossi, and Tomasi[28] and gas-phase geometries. The UAHF model is a set of rules, based on atomic number, connectivity, and charge, for determining radii of the spheres used to define the solute/solvent boundary. The rules are determined empirically so that they reproduce experimental solvation energies for small molecules. The electrostatic contribution to the solvation energy is calculated using the dielectric PCM (D-PCM)[12b] and the ICOMP = 4 charge normalization procedure,[29] while the dispersion–repulsion and cavitation contributions are calculated by the methods of Floris et al.[30] and Pierotti,[31] respectively (these are default options in the Gaussian 98 program),

$$\Delta G_s(X) = \Delta G_{elec}(X) + \Delta G_{cav}(X) + \Delta G_{disp-rep}(X) \quad (2)$$

The UAHF parametrization was done at the RHF/6-31G(d) level of theory for neutral molecules and cations and RHF/6-31+G(d) for anions, using gas-phase geometries. This study

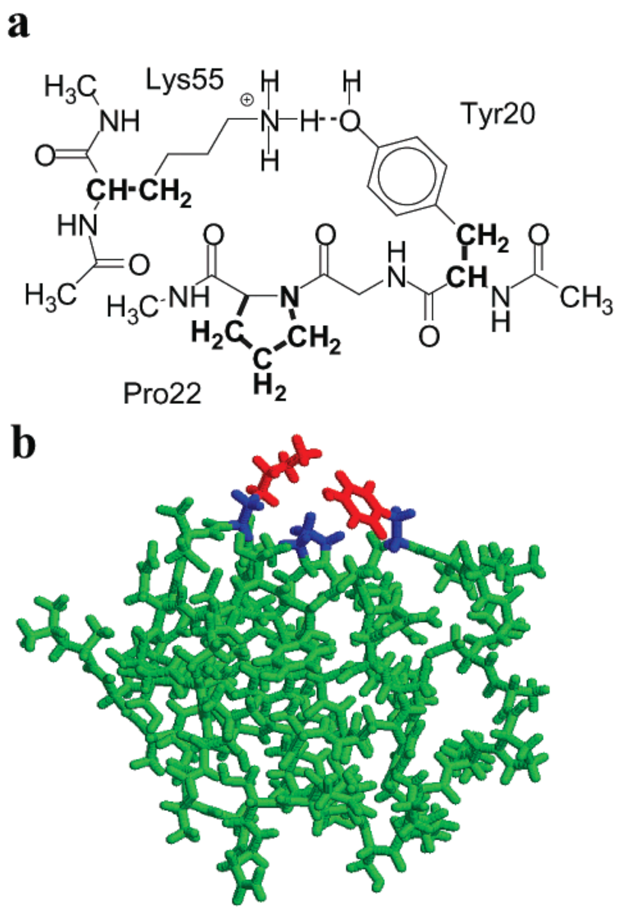


Figure 3. Subsystem of OMTKY3 (a) used to obtain the buffer region (bold) used for (b) ab initio/buffer/EFP regions (red/blue/green) used for the computation of the pK_a of Lys55.

uses the same level of theory for the calculation of the solvation energy, except that RHF/6-31+G(d)//RHF/6-31G(d) is used for anions to avoid optimizing geometries twice.

Because the ICOMP = 4 option is not available in the GAMESS code, we have also explored the effect of using the integral equation formalism[12c] PCM (IEF-PCM) without charge renormalization (ICOMP = 0) to calculate solvation energies. Following Topol et al.,[9] we use an experimental value of −262.5 kcal/mol for $\Delta G_s(H^+)$.

B. Protein pK_a Predictions. The solution structure of OMTKY3 has been determined using NMR by Hoogstraten et al.[32] and was obtained from the Protein Data Bank (entry 1OMU). We use the first of the 50 conformers without further refinement of the overall structure. The construction of the buffer and EFP regions used here for the calculation of the pK_a of Lys55 has been discussed in detail in a previous paper[33] and is only summarized here.

(1) The electronic and geometric structures of the Lys55 and Tyr20 side chains are treated quantum mechanically at the MP2/6-31+G(2d,p)//RHF/6-31G(d) level of theory (Figure 3), while the rest of the protein is treated with an effective fragment potential (EFP, described in more detail below). The use of the diffuse functions on atoms near the buffer region causes SCF convergence problems due to couplings with the induced dipoles in the EFP region, so the 6-31+G(2d,p) basis set was used only for the $C^{\delta}H_2C^{\epsilon}H_2NH_3\cdots HO-C^{\xi}(C^{\epsilon 1,2}H)_2$ atoms in the MP2 calculation.

(2) The ab initio region is separated from the protein EFP by a buffer region[34] comprising frozen localized molecular orbitals (LMOs) corresponding to all of the bond LMOs connecting the bold atoms in Figure 3a, as well as the core and lone pair LMOs belonging to those atoms. The Pro22 buffer is needed to describe its short-range interactions with Tyr20.[33] The buffer LMOs are generated by an RHF/6-31G(d) calculation on a subset of the system (shown in Figure 3a), projected onto the buffer atom basis functions[35] and subsequently frozen in the EFP calculations by setting select off-diagonal MO Fock matrix elements to zero.[36,37] The ab initio/buffer region interactions are calculated ab initio and thus include short-range interactions.

(3) The EFP describing the rest of the protein is generated by nine separate ab initio calculations on overlapping pieces of the protein truncated by methyl groups. Two different regions of overlap are used depending on whether it occurs on the protein backbone or on a disulfide bridge, as described in ref 34. The electrostatic potential of each protein piece is expanded in terms of multipoles through octupoles centered at all atomic and bond midpoint centers using Stone's distributed multipole analysis.[38] The monopoles of the entire EFP are scaled to ensure a net integer charge and the dipole polarizability tensor due to each LMO in the EFP region is calculated by a perturbation expression, as described in ref 34.

The EFP, buffer, and ab initio regions are combined and the geometry of the ab initio region is optimized at the RHF/6-31G(d) level of theory, while the geometry of the buffer and EFP regions are fixed. In a second calculation, a proton is removed from the amine group and the geometry of the ab initio region is reoptimized. Single-point energies ($E^{MP2//RHF}$) are evaluated at the MP2/6-31+G(2d,p) level of theory by excluding excitations from the buffer LMOs (and the core MOs in the ab initio region).[32]

Free Energy. The vibrational free energy (G^{vib}) of the optimized part of the ab initio region is calculated by the partial Hessian vibrational analysis (PHVA) method developed by Li and Jensen.[39] This method is based on a method by Head[40] in which only a subset of the atoms (in our case the atoms in the ab initio region) are displaced during a numerical Hessian calculation, to calculate a "partial Hessian". Further studies by Li and Jensen[39] have shown that vibrational energy and entropy changes for proton abstraction reactions calculated using frequencies obtained in this manner are within 0.2 kcal/mol of conventional values. The translational and rotational free energies (G^{trans} and G^{rot}) are calculated using the atomic masses and positions of all atoms in the protein.

Solvation Energy. The solvation energy (ΔG_s) is calculated using the ONIOM-PCM/X approach,[41] which combines IEF-PCM/ICOMP = 0 protein solvation energies with D-PCM/ICOMP = 4 solvation energies of model systems,

$$\begin{aligned}\Delta G_s(\text{protein,D-PCM/ICOMP=4}) = &\\ \Delta G_s(\text{protein,IEF-PCM/ICOMP=0}) + &\\ \Delta G_s(\text{model,D-PCM/ICOMP=4}) - &\\ \Delta G_s(\text{model,IEF-PCM/ICOMP=0}) & \quad (3)\end{aligned}$$

The model systems used will be discussed in section III. UAHF spheres are used for the atoms of the entire system, and the cavitation and dispersion−repulsion energies are calculated as above. The protein solvation energies are calculated using the EFP/IEF-PCM interface developed by Bandyopadhyay, Gordon, Mennucci, and Tomasi[26] and extended to protein-sized systems for the calculations described here by decreasing the memory requirement and parallelizing the code.

TABLE 1: Computed and Experimental pK_a's of Small Molecules with Functional Groups Found in Amino Acid Residues and the Individual Energy Components Used to Compute the pK_a's (Figure 2 and eq 1) in kcal/mol

			D-PCM/ICOMP = 4		IEF-PCM/ICOMP = 0		
acid	$\Delta E^{\mathrm{MP2}\ a}$	$\Delta G_{\mathrm{trv}}{}^{b}$	$\Delta\Delta G_s{}^{c,e}$	pK_a	$\Delta\Delta G_s{}^{c}$	pK_a	expt
acetic acid	352.55	−13.28	−333.08	4.6	−330.81	6.2	4.8
methylamine	223.28	−13.25	−196.72	9.8	−196.76	9.8	10.6
imidazole	231.26	−12.66	−210.34	6.1	−210.19	6.2	7.0
phenol	354.43	−11.99	−329.20	9.7	−322.88	14.4	10.0
methanethiol	360.71	−10.02	−336.90	10.1	−331.85	13.8	10.3
rmsd^d				0.6		2.6	
Lys55	249.38	−12.26	−221.78	11.3	−223.61	9.9	11.1

[a] Gas-phase (electronic) deprotonation energy, $\Delta E^{\mathrm{MP2//RHF}}$; cf. eq 1. [b] Gas-phase free energy correction using 1 M reference state; sum of the last four terms in eq 1. [c] Change in solvation energy; last three terms in the equation in Figure 2. [d] Root-mean-square deviation from experiment. [e] Calculated using D-PCM-X/ICOMP = 4; cf. eq 3.

Briefly, the isotropic IEF-PCM implementation involves the determination of the induced surface charges at each tesserae (**q**) through the equation

$$\mathbf{q} = \mathbf{D}^{-1}\mathbf{V} \tag{4}$$

where **V** is the solute electrostatic potential at each tesserae. The matrix $\mathbf{D}^{-1}$,

$$\begin{aligned}\mathbf{D}^{-1} &= \mathbf{A}^{-1}\left[\left(\frac{\epsilon+1}{\epsilon-1}\frac{\mathbf{A}}{2} - \mathbf{D}'\right)\mathbf{A}^{-1}\mathbf{S}\right]^{-1}\left(\frac{\mathbf{A}}{2} - \mathbf{D}'\right)\\ &= \mathbf{A}^{-1}\mathbf{T}^{-1}\mathbf{M}\\ &= \mathbf{A}^{-1}\mathbf{C}^{-1}\end{aligned} \tag{5}$$

is of dimension $N_{\mathrm{ts}}{}^2$, where N_{ts} is the number of tesserae, as are **S** and **D′**, while **A** is diagonal and only requires N_{ts} words of storage [see ref 12c for explicit expressions of their matrix elements]. Because N_{ts} is roughly 10 000 for OMTKY3, the manipulation and inversion of these matrices require significant CPU resources and memory ($3N_{\mathrm{ts}}{}^2$ or ca. 3 GB of RAM).

Here, the memory requirements were reduced to $N_{\mathrm{ts}}{}^2$ by implementing a "direct" construction of **T** and $\mathbf{C}^{-1}$ in which **S** and **D′** are not stored. The extra expense of recomputing these matrices is more than recouped by parallelizing the construction of **T** and $\mathbf{C}^{-1}$, which scale as $N_{\mathrm{ts}}{}^3$. The scaling with respect to the number of nodes (n) has been tested up to $n = 4$, and near-linear scaling was observed for these steps (note that the memory requirement is $nN_{\mathrm{ts}}{}^2$). The computation of $\mathbf{T}^{-1}$ is not easily parallelized, so the CPU time required to obtain **q** still scales as $N_{\mathrm{ts}}{}^3$. However, empirically, we find that the construction of **T**, $\mathbf{T}^{-1}$, and $\mathbf{C}^{-1}$ roughly scale as $4N_{\mathrm{ts}}{}^2$, $N_{\mathrm{ts}}{}^2$, and $3N_{\mathrm{ts}}{}^2$, respectively. Thus, the overall CPU time should be reduced roughly 8-fold for large n (assuming linear scaling past $n = 4$).

The EFP/PCM interface is similar to an all ab initio PCM calculation except that the electrostatic potential (**V**) of the EFP region is due to its multipole representation of the electrostatic potential. The induced surface charges influence the induced dipoles and this contribution is iterated to self-consistency. In this study we found several cases of divergence, presumably where surface charges are close to a polarizability tensor. Thus, the polarizability tensors are removed for the single-point calculations necessary for the solvation energies.

In the current implementation, $\Delta G_{\mathrm{disp-rep}}$ [cf. eq 2] is calculated only for the ab initio and buffer region. Furthermore, surface smoothing by the generation of additional spheres is prevented by using RET = 100 in the $PCM group, because the number of added spheres never converged for the protein within the memory available.

TABLE 2: Comparison of Computed pK_a's Presented Here to Previous Studies Where Applicable

acid	ref 8	ref 9	ref 11	ref 13	this work
acetic acid	4.9		8.2	4.3	4.6
methylamine	12.2				9.9
imidazole	8.7	6.9			6.2
phenol	10.7				9.8
methanethiol		10.4		5.8	10.2

Miscellaneous. The Foster−Boys localization procedure was used throughout this work to generate LMOs,[42] and all calculations were done with the quantum chemistry code GAMESS,[43] except the D-PCM/ICOMP = 4 calculations, which were done using Gaussian 98.[44]

III. Results and Discussion

A. Small Molecule pK_a Predictions. The predicted pK_a's for acetic acid, methylamine, imidazole, phenol, and methanethiol (small-molecule models of common ionizable groups in proteins) are listed in Table 1, together with the energy components defined in eqs 1 and 2. The D-PCM/ICOMP = 4 procedure results in pK_a's that are within 0.9 pH units of experiment with a root-mean-square deviation (rmsd) of 0.6 pH units. All predicted pK_a's are underestimated, and the rmsd can be decreased to 0.3 pH units by increasing $\Delta G_s(\mathrm{H}^+)$ to −261.8 kcal/mol, which is well within the range of experimental estimates.[9]

Table 2 offers a comparison to previous published methods where applicable. With the exception of the method by Richardson et al.,[8] a thorough comparison is difficult because a different group of molecules was considered. However, the data in Table 2 indicate that the method proposed here is at least as accurate as previously published methods and often more accurate.

The ICOMP = 4 charge renormalization method is not available in the GAMESS program, so the use of IEF-PCM/ICOMP = 0 (i.e., no charge renormalization) is investigated here. The ICOMP = 2 option (i.e., uniform charge renormalization), available in GAMESS, is not considered because that option leads to unphysical results for systems with several charged groups[27] such as proteins. The use of IEF-PCM/ICOMP = 0 for the calculation of the solvation energies leads to essentially unchanged results for methylamine and imidazole (Table 1). Larger errors are observed for acetic acid, phenol, and methanethiol because charge penetration into the continuum is more pronounced for anions. Thus, for the calculation of protein pK_a's, it will be necessary to estimate the effect of D-PCM/ICOMP = 4 on the ionizable residue in question. This is accomplished by the ONIOM-PCM/X method as described in section II.B and below.

B. Protein pK_a Predictions. *Single Protonation Site.* For a molecule with a single protonation site,

$$BH^+ \leftrightarrow B + H^+ \qquad K_a = \frac{[B][H^+]}{[BH^+]} \tag{6}$$

the relationship between the standard free energy of the reaction (ΔG_{rxn}) and the measured pK_a is straightforward (Figure 2). The pK_a can be measured by following a pH-dependent observable over a range of pH to determine the pH at which the protonation probability (θ) is 0.5,

$$\theta = \frac{10^{pK_a - pH}}{1 + 10^{pK_a - pH}} \tag{7}$$

Multiple Titratable Sites. In a protein, there are usually many (n) ionizable residues. The pK_a of a particular site (i) is still measured by determining a titration curve and finding the pH for which $\theta_i = 0.5$ [cf. eq 7]. However, θ_i is in principle determined by the pK_a's of all ionizable residues in each possible protonation state,[2e]

$$\theta_i = \langle x_i \rangle = \frac{\sum_j^{2^n} x_i^j\, e^{-G_j/(RT)}}{\sum_j^{2^n} e^{-G_j/(RT)}} \tag{8}$$

(here x_i^j is 1 or 0 depending on whether site i is protonated or unprotonated in protonation state j, respectively. G_j is the free energy of protonation state j at a given pH). Thus, the resulting apparent pK_a of site i (p$K_{app,i}$) must, in principle, be obtained by computing the free energies for all sites for all possible (2^n) protonation states. Clearly, this not feasible with the QM/MM/LPBE method because of the computational expense of the QM component.

To proceed, we consider the error incurred by using a single ΔG_{rxn}, that is, a single ionization state, to compute the apparent pK_a for a simple system with two cationic sites ($^+HB_1B_2H^+$) as discussed by Bashford and Karplus.[5a] The intrinsic pK_a (pK_{intr}) of each group is the pK_a of a group when the other group is neutral.

$$^+HB_1B_2 \leftrightarrow B_1B_2 + H^+ \qquad pK_{intr,1}$$

$$B_1B_2H^+ \leftrightarrow B_1B_2 + H^+ \qquad pK_{intr,2}$$

When group 2 is charged the pK_{intr} of group 1 is shifted by the interaction energy between the two charged groups (W, in pH units), and *vice versa* for p$K_{intr,2}$.

$$^+HB_1B_2H^+ \leftrightarrow B_1B_2H^+ + H^+ \qquad pK'_{intr,1} = pK_{intr,1} - W$$

$$^+HB_1B_2H^+ \leftrightarrow {}^+HB_1B_2 + H^+ \qquad pK'_{intr,2} = pK_{intr,2} - W$$

The protonation probability of site 1 is[5a]

$$\theta_1 = \frac{10^{pK_{intr,1} - pH} + 10^{pK_{intr,1} + pK_{intr,2} - 2pH - W}}{1 + 10^{pK_{intr,1} - pH} + 10^{pK_{intr,2} - pH} + 10^{pK_{intr,1} + pK_{intr,2} - 2pH - W}} \tag{9}$$

The difference between p$K_{app,1}$ and p$K_{intr,1}$ is a complicated function of the difference in the intrinsic pK_a's of site 1 and 2 and the magnitude of W.

However, Bashford and Karplus have shown that for p$K_{intr,1} \neq$ p$K_{intr,2}$ the mean-field approximation due to Tanford and Roxby,[47]

$$pK_{app,1} = pK_{intr,1} - \theta_2 W \tag{10}$$

is a good approximation even for strongly coupled sites. Thus, if θ_2 and W are known, the error incurred by approximating θ_2 by either 0 or 1 can easily be estimated.

In the case of p$K_{intr,1}$ = p$K_{intr,2}$, the mean-field approximation breaks down for large W and eq 10 may no longer be a valid approximation. A new relation between p$K_{app,1}$ and p$K_{intr,1}$ can be obtained by an analysis of eq 9 for the case p$K_{intr,1}$ = p$K_{intr,2}$,

$$\theta_1 = \frac{10^{pK_{intr,1} - pH} + 10^{2pK_{intr,1} - 2pH - W}}{1 + 2(10^{pK_{intr,1} - pH}) + 10^{2pK_{intr,1} - 2pH - W}} \tag{11}$$

The apparent pK_a's can be obtained by finding the inflection points

$$\frac{\partial^2 \theta_1}{\partial (pH)^2} = 0 \tag{12}$$

on the titration curve. The resulting equation has three solutions,

$$pK^0 = pK_{intr,1} - \frac{1}{2}W$$

$$pK^{\pm} = pK_{intr,1} + \log[1 - 10^{-W}(2 \pm \sqrt{4 - 5 \times 10^W + 10^{2W}})] \tag{13}$$

The dependencies of these pK's on W are illustrated in Figure 5. For $W \leq \log(4)$, p$K^{\pm}$ is complex and Re(p$K^{\pm}$) = pK^0, that is, there is a single inflection point on the titration curve and the corresponding p$K_{app,1}$ is the mean field result of eq 10 (Figure 4a).

For $W > \log(4)$, pK^+ and pK^- approach the following limits for large W:

$$pK^+ = pK_{intr,1} + \log(2) \tag{14a}$$

$$pK^- = pK_{intr,1} - W - \log(2) = pK'_{intr1} - \log(2) \tag{14b}$$

The factor of log(2) reflects the gain or loss in entropy due to 2-fold energy degeneracy in the products and reactants, respectively (Figure 4b). Moreover, the change in the titration curve on going from $W \leq \log(4)$ to $W > \log(4)$ shown in Figure 4 is a result of W becoming larger than the entropy gain due to 4-fold degeneracy of the products [log(4)] inherent in the mean field approximation.

For a system with two anionic sites (HA_1A_2H), a similar analysis yields

$$pK^0 = pK_{intr,1} + \frac{1}{2}W$$

$$pK^{\pm} = pK_{intr,1} - \log[1 - 10^{-W}(2 \pm \sqrt{4 - 5 \times 10^W + 10^{2W}})] \tag{15}$$

for p$K_{intr,1}$ = p$K_{intr,2}$.

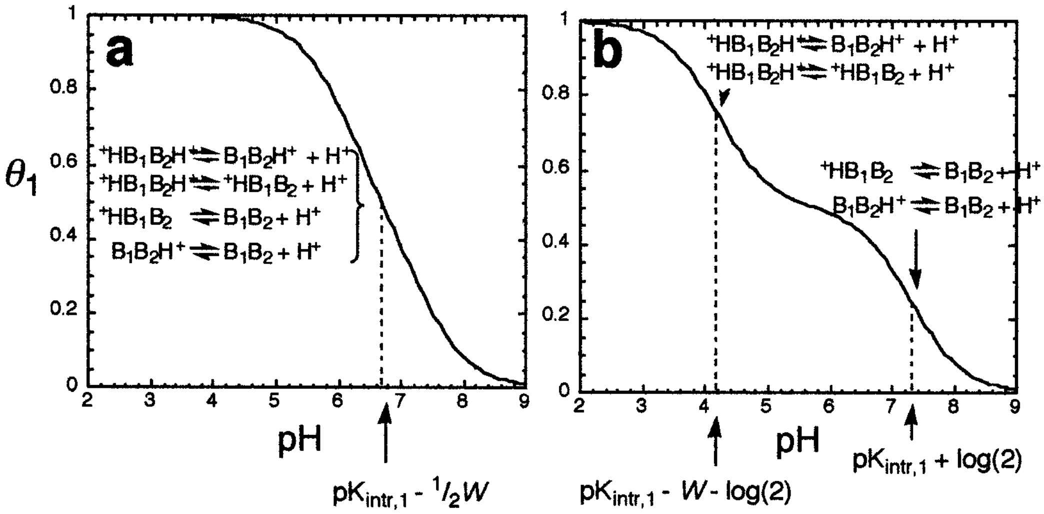


Figure 4. θ_1 (eq 9) vs pH for $pK_{intr,1} = pK_{intr,2} = 7.0$ pH units for (a) $W = 0.60$ pH units and (b) $W = 2.5$ pH units.

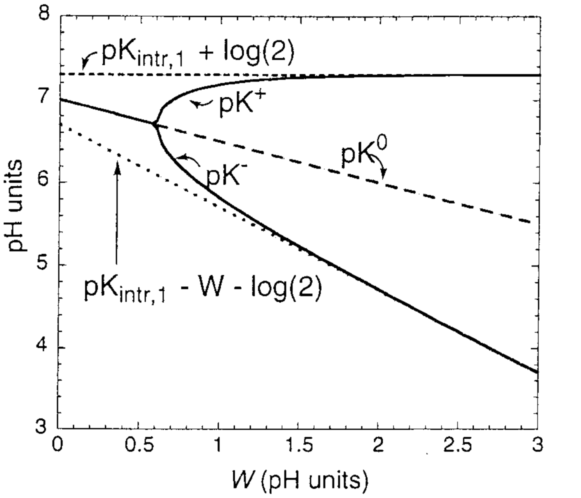


Figure 5. Plot of the apparent pK's of eq 13 vs W. For $W \leq \log(2)$, pK^+ and pK^- are complex numbers and only the real parts are plotted. Also plotted are the limits for pK^+ and pK^- as W increases (eq 14).

For a system with one anionic and cationic site ($HA_1B_2H^+$), sites 1 and 2 can have equal apparent pK_a's when $pK_{intr,1} - W = pK_{intr,2}$,

$$HA_1B_2H^+ \leftrightarrow {}^-A_1B_2H^+ + H^+ \qquad pK'_{intr,1} = pK_{intr,1} - W$$

$$HA_1B_2H^+ \leftrightarrow HA_1B_2 + H^+ \qquad pK_{intr,2}$$

The protonation probabilities for this system are

$$\theta_1 = \frac{10^{pK_{intr,1}-pH} + 10^{pK_{intr,1}+pK_{intr,2}-2pH}}{1 + 10^{pK_{intr,1}-pH} + 10^{pK_{intr,2}-pH+W} + 10^{pK_{intr,1}+pK_{intr,2}-2pH}} \qquad (16a)$$

$$\theta_2 = \frac{10^{pK_{intr,2}-pH+W} + 10^{pK_{intr,1}+pK_{intr,2}-2pH}}{1 + 10^{pK_{intr,1}-pH} + 10^{pK_{intr,2}-pH+W} + 10^{pK_{intr,1}+pK_{intr,2}-2pH}} \qquad (16b)$$

However, for $pK_{intr,1} - W = pK_{intr,2}$, both equations reduce to eq 11, and the apparent pK_a's are given by eqs 12 and 13, that is, $pK_{app,1} = pK_{app,2} = pK^0$ or pK^- for $W \leq \log(4)$ and $W > \log(4)$, respectively.

The analysis presented here is a special case of the decoupled site representation (DSR) of Onufriev, Case, and Ullmann.[45] The DSR is a numerical technique for extracting the pK_a's of individual "quasi-sites" from titration curves of interacting sites. The pK_a's of eqs 13 and 14 represent quasi-site pK_a's for a two-site system with identical pK_a's. For example, quasi-site pK_a's of a titration curve for two cationic sites with pK_{intr} values of 7.0 and 7.1 and $W = 2.2$ pH units (3 kcal/mol) are predicted to be 4.6 and 7.4 by the DSR method. These values are reproduced quite well (4.5 and 7.4) by eq 14 despite the fact that it was derived for $pK_{intr,1} = pK_{intr,2}$. This agreement is encouraging because nearly identical pK_{intr}'s will be more common than truly identical pK_{intr}'s in most chemical systems.

The conclusions reached from this analysis of a system with two titratable sites regarding the use of a single protonation state for the calculation of the apparent pK_a can be summarized as follows:

(1) $^+HB_1B_2H^+$

A. $pK_{intr,1} - pK_{intr,2} < 0$

$$pK_{app,1} \approx pK'_{intr,1}; \quad \text{error} \approx (1 - \theta_2^1)W$$

$$pK_{app,2} \approx pK_{intr,2}; \quad \text{error} \approx \theta_1^2 W$$

B. $pK_{intr,1} - pK_{intr,2} > 0$

$$pK_{app,1} \approx pK_{intr,1}; \quad \text{error} \approx \theta_2^1 W$$

$$pK_{app,2} \approx pK'_{intr,2}; \quad \text{error} \approx (1 - \theta_1^2)W$$

C. $pK_{intr,1} = pK_{intr,2}, \quad W \leq \log(4)$

$$pK_{app} \approx pK_{intr,1} \approx pK'_{intr,1} \approx pK_{intr,2} \approx pK_{intr,2}; \quad \text{error} \approx {}^1/_2W$$

D. $pK_{intr,1} = pK_{intr,2}, \quad W > \log(4)$

$$pK_{app,a} \approx pK'_{intr,1} - \log(2) \approx pK'_{intr,2} - \log(2); \quad \text{error} = (1 - \theta_b^a)W$$

$$pK_{app,b} \approx pK_{intr,1} + \log(2) \approx pK_{intr,2} + \log(2); \quad \text{error} = \theta_a^b W$$

(2) HA_1A_2H

A. $pK_{intr,1} - pK_{intr,2} < 0$

$$pK_{app,1} \approx pK_{intr,1}; \quad \text{error} \approx \theta_2^1 W$$

$$pK_{app,2} \approx pK'_{intr,2}; \quad \text{error} \approx (1 - \theta_1^2)W$$

B. $pK_{intr,1} - pK_{intr,2} > 0$

$$pK_{app,1} \approx pK'_{intr,1}; \quad \text{error} \approx (1 - \theta_2^1)W$$

$$pK_{app,2} \approx pK_{intr,2}; \quad \text{error} \approx \theta_1^2 W$$

C. $pK_{intr,1} = pK_{intr,2}, \quad W \leq \log(4)$

$$pK_{app} \approx pK_{intr,1} \approx pK'_{intr,1} \approx pK_{intr,2} \approx pK'_{intr,2}; \quad \text{error} \approx {}^1/_2 W$$

D. $pK_{intr,1} = pK_{intr,2}, \quad W > \log(4)$

$$pK_{app,a} \approx pK_{intr,1} - \log(2) \approx pK_{intr,2} - \log(2); \quad \text{error} \approx \theta_b^a W$$

$$pK_{app,b} \approx pK'_{intr,1} + \log(2) \approx pK'_{intr,2} + \log(2); \quad \text{error} \approx (1 - \theta_a^b)W$$

(3) $HA_1B_2H^+$

A. $pK'_{intr,1} - pK_{intr,2} < 0$

$$pK_{app,1} \approx pK'_{intr,1}; \quad \text{error} \approx (1 - \theta_2^1)W$$

$$pK_{app,2} \approx pK'_{intr,2}; \quad \text{error} \approx \theta_1^2 W$$

B. $pK'_{intr,1} - pK_{intr,2} > 0$

$$pK_{app,1} \approx pK_{intr,1}; \quad \text{error} \approx \theta_2^1 W$$

$$pK_{app,2} \approx pK_{intr,2}; \quad \text{error} \approx (1 - \theta_1^2)W$$

C. $pK'_{intr,1} = pK_{intr,2}, \quad W \leq \log(4)$

$$pK_{app} \approx pK_{intr,1} \approx pK'_{intr,1} \approx pK_{intr,2} \approx pK'_{intr,2}; \quad \text{error} \approx {}^1/_2 W$$

D. $pK'_{intr,1} = pK_{intr,2}, \quad W > \log(4)$

$$pK_{app,a} \approx pK'_{intr,1} - \log(2) \approx pK_{intr,2} - \log(2); \quad \text{error} \approx (1 - \theta_b^a)W$$

$$pK_{app,b} \approx pK_{intr,1} + \log(2) \approx pK'_{intr,2} + \log(2); \quad \text{error} \approx \theta_a^b W$$

(4) $W = 0$

$$pK_{app,1} \approx pK_{intr,1} \approx pK'_{intr,1}; \quad \text{error} \approx 0$$

$$pK_{app,2} \approx pK_{intr,2} \approx pK'_{intr,2}; \quad \text{error} \approx 0$$

Here, θ_x^y is eq 7 evaluated for pH = $pK_{app,y}$ and $pK_a = pK_{app,x}$, and a and b refer a quasi-site (two groups in this case) rather than one particular group. These conclusions also apply to systems with more than two ionizable sites provided the pK_a's of invidual sites are only determined by pairwise interactions.

The general conclusion reached here is in complete accord with Gilson's[47] finding that "the free energy of a single highly populated ionization state will frequently represent an excellent estimate of the overall ionization energy", which forms the basis for his predominant state approximation.

TABLE 3: Experimentally Measured pK_a's and Hill Coefficients (cf. eq 12) for Ovomucoid Third Domain, Taken from Ref 24[a]

residue	exptl pK_a	n	calcd pK_a
Asp27	<2.3	0.85	3.4
CysC56	<2.5	0.87	2.7
Asp7	<2.7	0.72	3.3
Gly19	3.2	1.08	2.8
Glu10	4.2	0.92	3.5
Glu43	4.8	0.95	4.4
His52	7.5	0.93	6.2
LeuN1	8.0	0.88	7.5
Lys13	9.9	0.69	11.2
Lys34	10.1	0.66	11.7
Tyr11	10.2	0.73	10.0
Tyr20	11.1	0.57	9.9
Lys29	11.1	0.87	12.1
Lys55	11.1	0.64	11.3
Tyr31	>12.5		11.2

[a] Standard pK_a values for amino acid residues are as follows: Asp = 4.0; Glu = 4.4; Tyr = 9.6; His = 6.6−7.0; Lys = 10.4; α-carboxyl group = 3.8; α-amino group = 7.5

The initial intent of the QM/MM/LPBE method described above is to rationalize unusual pK_a's that have already been identified experimentally and for which MM/LPBE pK_a predictions already have been (or easily could be) performed. Thus, the calculated or measured pK_a's and site−site interactions can be used to determine the optimum protonation state at a given pH. Furthermore, the MM/LPBE results provide the interactions between sites that can be used to estimate the error incurred by using a single protonation state of the protein to calculate the pK_a of a particular site, as demonstrated next.

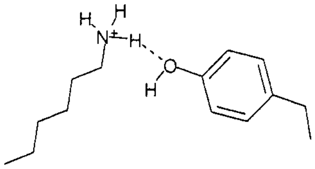
The pK_a of Lys55. Table 3 lists the experimentally and computationally determined pK_a's of OMTKY3,[23,24] which together with MM/LPBE predictions[24] of the interaction between sites (W) can be used to determine the optimum protonation state of each ionizable residue for the calculation of the pK_a of Lys55.

The MM/LPBE calculations predict that the pK_a of Lys55 is affected appreciably (i.e., $|W| > 0.1$ pH units) by only three residues, Tyr20, His52, and CysC56, so that any protonation state can be used for the remaining residues. The respective experimental pK_a's of His52 and CysC56 residues are >8.6 and 3.6 pH units lower than that of Lys55. Both are therefore >99.999% deprotonated at pH = 11.1 and can be treated as 100% deprotonated for the calculation of the pK_a of Lys55. A similar conclusion is reached by using the apparent pK_a's from the MM/LPBE calculations.

Finally, the experimental pK_a's of Tyr20 and Lys55 are equal and predicted to interact by $W = 0.6$ pH units. Because W is very close to log(4), either case 3C or case 3D discussed in the previous section should apply. Both cases suggest that unionized Tyr20 provides an appropriate protonation state for the calculation of the pK_a of Lys55, provided that the calculated pK_a is reduced by 0.3 [log(2) or ${}^1/_2W$ in case 3C or D, respectively]. We note that the MM/LPBE approach underestimates the pK_a of Tyr20 and suggests that ionized Tyr20 is the optimum protonation state for the calculation of the pK_a of Lys55. This error can have two sources: errors in the intrinsic pK_a or a wrong W, or both. The most likely error is an underestimation of the pK_{intr} of Tyr20.[48] In this case, the error incurred by using the ionized state of Tyr20 for the prediction of the pK_a of Lys55 should be $(1 - {}^1/_2)W$ or 0.3 pH units.

The MM/LPBE prediction that the pK_a's of Lys55 and Tyr20 are coupled (i.e., that W is nonzero) is supported by the similar Hill coefficients ($n = 0.64$ and 0.57) for Lys55 and Tyr20. The

SCHEME 1



Hill coefficient (n) is a measure of the deviation from a standard sigmoidal titration curve [cf. eq 7],

$$\theta_i = \frac{10^{n(\mathrm{p}K_{\mathrm{a,i}}-\mathrm{pH})}}{1 + 10^{n(\mathrm{p}K_{\mathrm{a,i}}-\mathrm{pH})}}$$

and deviation from 1 is usually taken as a measure of coupling between titration sites. We note that $n = 0.87$ for Lys29 (the pK_a of which is equal to that of Lys55), although a Hill coefficient close to 1 does not always imply that the deprotonation of the residue is uncoupled from other groups.[46]

Thus, on the basis of previous experimental and computational results, the use of the pH = 7 protonation state used in the NMR refinement (Asp, Glu = negative; His, Tyr = neutral; Arg, all other Lys = positive) to compute the pK_a of Lys55 should lead to a computed pK_a that is 0.3 pH units too high.[24]

The computed pK_a of Lys55 is 11.3 (Table 1) which, when reduced by 0.3 pH units, is in excellent agreement with the experimental value of 11.1, and within the estimated experimental uncertainty of 0.2 pH units.[24] This result suggests that the MM/LPBE results present a sufficiently accurate picture of interresidue interactions for the determination of a suitable protonation state for QM/MM/LPBE predictions of pK_a's. However, further studies using different protonation states are required to verify these results.

The solvation energies are calculated using eq 3 and a model system consisting of the side chains of Lys55 and Tyr20 (shown in Scheme 1 for the protonated state). Interestingly, the correction to the IEF-PCM/ICOMP = 0 solvation energy change for Lys55 is −1.83 kcal/mol, which is substantially larger than that for methylamine (Table 1). This difference is presumably due to the hydrogen bond between Lys55 and Tyr20.

In our current model, deprotonation of Tyr20 results in a *spontaneous* proton transfer from Lys55 to Tyr20, to yield the same $-\mathrm{OH}\cdots\mathrm{NH}_2-$ hydrogen-bonded structure that resulted from Lys55 deprotonation. Thus, the pK_a's of Lys55 and Tyr20 are coupled because of the intramolecular hydrogen bond between the two residues, and the pK_a of Tyr20 is predicted to be equal to that of Lys55 according to the method outlined in Figure 2. This is in agreement with experiment, as shown in Table 3. However, the proton transfer may be an artifact of the use of geometry optimizations in the gas phase, where the relative basicities are significantly different from those in solution. We are currently developing a slightly modified computational methodology for pK_a predictions based on geometry optimizations using PCM to address this issue and will present our findings in a future paper.

IV. Summary and Future Directions

This paper presents a computational methodology for pK_a predictions of small molecules based on an ab initio quantum mechanics (QM) description of the acid and a linearized Poisson−Boltzmann equation (LPBE) description of bulk solvation. The method is very similar to other QM/LPBE methods for small molecule pK_a predictions[6−11,13,14] but uses Tomasi's polarized continuum method (PCM)[12] in conjunction with the united atom-Hartree−Fock radii proposed by Barone, Cossi, and Tomasi.[26] This QM/LPBE method is capable of reproducing the pK_a's of several functional groups found in amino acid residues with a root-mean-square deviation from experiment of 0.6 pH units (Table 1).

The practical applicability of the QM/LPBE method is extended to proteins by using a QM description of the ionizable residue and an effective fragment potential (EFP, a hybrid QM/MM method[25]) description of the rest of the protein. This QM/MM/LPBE methodology for protein pK_a predictions is made possible by the recent interface between the EFP and PCM methods, due to Bandyopadhyay, Gordon, Mennucci, and Tomasi.[26] The method is used to predict a pK_a of the Lys55 residue in turkey ovomucoid third domain that is in good agreement with experiment:[24] 11.0 vs 11.1.

Clearly this study presents only the first, proof-of-concept step in the development of general computational methods for protein pK_a predictions based on QM/MM. The following issues are currently being investigated or will be the studied soon.

The protein pK_a results presented in this study are obtained by using a single protonation state corresponding to neutral pH. The use of other protonation states and a priori methods for their determination will be investigated, such as using pK_{model} values and a pH slightly higher or lower than the pK_{model} value of the residue of interest.

We are exploring the use of geometry optimizations using PCM for systems in which proton transfer within the ab initio region is possible. Future studies will focus on the use of molecular dynamics simulations to better sample the conformational space of the ab initio region and the rest of the protein.

The pK_a of Lys55 is shifted relatively modestly compared to the solution-phase values. The prediction of larger pK_a shifts may provide a more rigorous test of our computational method.[4] Calculations of the pK_a's of Glu35 in lysozyme, Asp25 in HIV, and Glu172 in Xylanase (all with pK_a increases of nearly 3 pH units relative to pK_{model}) are planned for the near future.

The calculation of the solvation energy currently represents a considerable computational bottleneck, both in terms of CPU time and memory requirements. For OMTKY3, roughly 10 days of CPU and 4 GB of RAM are required for the evaluation of the two solvation energies needed for a pK_a prediction using three nodes on a four-node RS/6000 44P 270 workstation. However, we are working on increasing the efficiency of the QM/MM/LBFE method by implementing a linear scaling PCM method[49] (in collaboration with Pomelli and Tomasi). In addition, we will explore the use of the surface generalized Born (SGB) method[49] for the calculation of the solvation energy by an ONIOM-PCM/GSB approach. These methodological improvements will allow for more routine evaluations of pK_a's in proteins that are larger than OMTKY3.

It is our hope that the QM/MM/LPBE results obtained here and in future studies will aid in the interpretation of experiment and guide the improvements of computationally more efficient methods and QM/MM prediction methods with an explicit representation of the solvent.[21,22]

Acknowledgment. The authors thank Michael Gilson for many very helpful discussions and Adrian Elcock for a careful critique of the manuscript. This work was supported by a Research Innovation Award from the Research Corporation, a type G starter grant from the Petroleum Research Fund, and a Carver Scientific Research Initiative Grant. H.L. gratefully acknowledges a predoctoral fellowship from the Center for Biocatalysis and Bioprocessing at the University of Iowa. The calculations were performed on IBM RS/6000 workstations obtained through a CRIF grant from the NSF (Grant CHE-9974502).

References and Notes

(1) Perrin, D. D.; Dempsey, B.; Serjeant, E. P. *pK_a Prediction for Organic Acids and Bases*; Chapman and Hall: London, 1981.

(2) (a) Warshel, A. *Biochemistry* **1981**, *20*, 3615. (b) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219. (c) Yang, A. S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins* **1993**, *15*, 252. (d) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol*. **1994**, *238*, 415. (e) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819. (f) For a review, see: Ullmann, G. M.; Knapp, E.-W. *Eur. Biophys. J.* **1999**, *28*, 533.

(3) (a) Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K.; McCammon, J. A. *J. Comput. Chem*. **1996**, *17*, 1633. (b) Demchuck, E.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 17373.

(4) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400.

(5) (a) Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9556. (b) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5804.

(6) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.

(7) Chen, J. L.; Noodleman, L.; Case, D. A.; Bashford, D. *J. Phys. Chem.* **1994**, *98*, 11059.

(8) Richardson, W. H.; Peng, C.; Bashford, D.; Noodleman, L.; Case, D. A. *Int. J. Quantum Chem*. **1997**, *61*, 207.

(9) (a) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Phys. Chem. A* **1997**, *101*, 10075. (b) Topol, I. A.; Burt, S. K.; Rashin A. A.; Erickson, J. W. *J. Phys. Chem. A* **2000**, *104*, 866. (c) Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; Burt, S. K. *J. Phys. Chem. A* **2000**, *104*, 9619. (d) Topol, I. A.; Nemukhin, A. V.; Dobrogorskaya, Y. A.; Burt, S. K. *J. Phys. Chem. B* **2001**, *105*, 11341.

(10) Jang, Y. H.; Sowers, L. C.; Cagin, R.; Goddard, W. A., III. *J. Phys. Chem. A* **2001**, *105*, 274.

(11) Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.

(12) (a) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys*. **1981**, *55*, 117. (b) Cossi, M.; Barone, V.; Cammi, R.; Tomasi. J. *Chem. Phys. Lett*. **1996**, *225*, 327. (c) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys*. **1997**, *107*, 3032.

(13) (a) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194. (b) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.

(14) (a) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc*. **2001**, *123*, 7314. (b) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys*. **2001**, *114*, 4595.

(15) Warshel, A.; Levitt, M. *J. Mol. Biol*. **1976**, *103*, 227.

(16) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5827.

(17) See “methods sections” of the following recent papers: (a) Alhambra, C.; Gao, J. *J. Comput. Chem*. **2000**, *13*, 1192. (b) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem*. **2000**, *21*, 1442. (c) Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc*. **2000**, *122*, 6560. (d) Cui, Q.; Karplus, M. *J. Am. Chem. Soc*. **2001**, *123*, 2284. (e) Billeter, S. R.; Hanser, C. F. W.; Mordasini, T. Z.; Scholten, M.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys*. **2001**, *3*, 688. (f) Turner, A. J.; Williams, I. H. *Phys. Chem. Chem. Phys*. **1999**, *1*, 1323. (g) Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Burton, N.; Hiller, I. A. *Theor. Chem. Acc*. **2001**, *106*, 105. (h) Antonczak, S.; Monard, G.; Ruiz-Lopez, M.; Rivail, J.-L. *J. Mol. Model*. **2000**, *6*, 527. (i) Bentzien, J.; Muller, R. P.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293.

(18) Singh, U. C.; Kollman, J. *Comput. Chem*. **1986**, *7*, 718.

(19) Thery, V.; Rinaldi, D.; J.-L.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. J. *Comput. Chem*. **1994**, *15*, 269.

(20) (a) Warshel, A.; Russell, S. T. Q. *Rev. Biophys*. **1984**, *17*, 283. (b) Warshel, A.; Russell, S. T. *J. Am. Chem. Soc.* **1986**, *108*, 6569. (c) Warshel, A.; Narayszabo, G.; Sussman, F.; Hwang, J. K. *Biochemistry* **1989**, *28*, 3629.

(21) Bash, P. A.; Ho, L. L.; MacKerell, A. D., Jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698.

(22) Byun, Y.; Mo, Y. R.; Gao, J. L. *J. Am. Chem. Soc*. **2001**, *123*, 3974.

(23) Schaller, W. *Robertson Biochem.* **1995**, *34*, 4714.

(24) Forsyth, W. R.; Gilson, W. R.; Antosiewicz, J.; Jaren, O. R.; Robertson, A. D. *Biochemistry* **1998**, *37*, 8643.

(25) (a) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293. (b) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Kraus, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

(26) Bandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.*, in press.

(27) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys*. **1972**, *56*, 2257.

(28) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys*. **1997**, *107*, 3210−3221.

(29) (a) Mennucci, B.; Tomasi, J. *J. Chem. Phys*. **1997**, *106*, 5151. (b) The electronic charge density extends into the continuum. Thus, the induced PCM charges arise from a noninteger molecular charge. The ICOMP = 4 procedure scales each induced charge by a factor-related value of the electron density at each tesserae center.

(30) Floris, F. M.; Tomasi, J.; Pascual Ahuir, J. L. *J. Comput. Chem*. **1991**, *12*, 784−791. The necessary parameters are taken from: Caillet, J.; Claverie, P.; *Acta Crystallogr*. **1978**, *B34*, 3266−3272.

(31) Pierotti, R. A. *Chem. Rev*. **1976**, *76*, 717.

(32) Hoogstraten, C. G.; Choe, S.; Westler, W. M.; Markley, J. L. *Protein Sci*. **1995**, *4*, 2289.

(33) Minikis, R. M.; Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2001**, *105*, 3829.

(34) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.

(35) King, H. F.; Stanton, R. E.; King, H.; Wyatt, R. E.; Parr R. G. *J. Chem. Phys.* **1967**, *47*, 1936.

(36) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15.

(37) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W., Jr. *J. Chem. Phys*. **1984**, *80*, 4378.

(38) Stone, A. J. *Chem. Phys. Lett.* **1981**, 83, 233.

(39) Li, H.; Jensen, J. H. *Theor. Chem. Acc*., in press.

(40) Head, J. D. *Int. J. Quantum Chem.* **1997**, *65*, 827−838.

(41) Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62.

(42) (a) Boys, S. F. In *Quantum Science of Atoms, Molecules and Solids*; Lowdin, P. O., Ed.; Academic Press: New York, 1966. (b) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.

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

(44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.

(45) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192.

(46) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413.

(47) Gilson, M. K. *Proteins* **1993**, *15*, 266.

(48) If the pK_{intr} of Tyr20 is correct, Lys55 must *raise* the pK_a of Tyr20 by 0.6 kcal/mol, that is, the interaction between C−O$^-$ and NH$_3^+$ would have to be repulsive.

(49) (a) Rega, N.; Cossi, M.; Barone, V. *Chem. Phys. Lett*. **1998**, *293*, 221. (b) Pomelli, C. S.; Tomasi, J. *J. Mol. Struct. (THEOCHEM)* **2001**, *537*, 97.

(50) Gosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.