pK_a Shifts in Small Molecules and HIV Protease: Electrostatics and Conformation

Rui Luo,[†,‡] Martha S. Head,[†,§,¶] John Moult,[†] and Michael K. Gilson*[,†,§]

Contribution from the Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850, Department of Chemistry and Biochemistry, University of Maryland at College Park, College Park, Maryland 20742, and National Institute of Standards and Technology, Gaithersburg, Maryland 20899

Received December 22, 1997. Revised Manuscript Received April 20, 1998

Abstract: The generalized Born (GB) approximation is a reasonable electrostatic model that is fast enough for use with extensive conformational sampling. This study combines the GB model with a torsion-space sampling method to compute pK_a shifts for a series of dicarboxylic acids and amino acids, and for the active-site aspartyl dyad in HIV-1 protease. The calculations agree rather well with experiment for the small molecules. Conformational analysis shows salt-bridging for the zwitterionic amino acids but otherwise modest electrostatic effects upon mean chain lengths. The calculations also show that through-space electrical fields alone cannot account completely for the observed pK_a shifts. The calculations for HIV protease agree reasonably well with experiment, despite the complexity of the system. The present computational approach should be useful for a variety of other applications.

Introduction

Classical electrostatic models—models based upon the Poisson equation or the Poisson−Boltzmann equation[1]—have long been used to compute the properties of molecules in solution. Before the advent of fast computers, the applicability of electrostatic models was limited to molecules whose simple shape made them amenable to analytic solutions. During the past decade, fast computers and numerical methods have allowed classical electrostatics to be applied to complex molecules (see, e.g., refs 2−7). However, the range of problems that can be addressed with such methods has been restricted by the computer time required to solve the electrostatic equations numerically. In particular, it has been difficult to carry out calculations that involve solving a new electrostatics problem for each of many molecular conformations.

Some years ago, the Generalized Born (GB) approximation was proposed as a computationally rapid replacement for more demanding solutions of the Poisson equation.[8] Although the original version was useful, it included an inefficient numerical step. Recently, more efficient versions of the GB approximation have been described.[9,10] These allow conformational sampling to be combined with a physically reasonable treatment of electrostatics, increasing the range of problems and systems that may be studied with continuum electrostatics.[10−13]

The present study combines the GB approximation with extensive conformational sampling to examine a series of systems for which experimental data exist. The method of conformational analysis that is used here, "mining minima" (MM), is a new one that directly yields conformational free energies for systems of modest size.[14] A closely related method has recently been described.[12] The combination of MM and GB to compute molecular properties is novel and the combination of these methods, if successful, would have a range of applications. The problems that are addressed here are as follows.

pK_a Shifts in Difunctional Compounds. Predicting the influence of polar substituents upon the pK_as of ionizable groups in small molecules is a classical problem in physical organic chemistry. The literature on this subject discusses two mechanisms by which polar substituents can influence pK_as: through-bond induction, and through-space electrical fields (see, e.g., refs 15 and 16). The electrostatic theory dates at least to 1923,[17−19] and a specific analytical model, that of Kirkwood

* Corresponding author. Phone: (301) 738-6217. FAX: (301) 738-6255. E-mail: gilson@indigo14.carb.nist.gov.
† University of Maryland Biotechnology Institute.
‡ University of Maryland at College Park.
§ National Institute of Standards and Technology.
¶ Current Address: SmithKline Beecham Pharmaceuticals, UW 2950, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939.

(1) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1973.
(2) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327−335.
(3) Gilson, M. K.; Honig, B. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 7−18.
(4) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(5) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem* **1993**, *97*, 1101−1109.
(6) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.
(7) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. *Rev. Comput. Chem.* **1994**, *5*, 229−267.
(8) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(9) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.* **1997**, *101*, 3005−3014.
(10) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(11) McDonald, D. Q.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 11550−11553.
(12) Kolossvary, I. *J. Phys. Chem. A* **1997**, *101*, 9900−9905.
(13) Kolossvary, I. *J. Am. Chem. Soc.* **1997**, *119*, 10233−10234.
(14) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem.* **1997**, *101*, 1609−1618.
(15) Topsom, R. D. *Prog. Phys. Org. Chem.* **1976**, *12*, 1−20.
(16) Reynolds, W. F. *Prog. Phys. Org. Chem.* **1983**, *14*, 165−203.
(17) Bjerrum, N. *Z. Phys. Chem.* **1923**, *106*, 219.
(18) Eucken, A. *Z. Angew. Chem.* **1932**, *45*, 203.

S0002-7863(97)04307-2 CCC: $15.00
Published on Web 05/29/1998

and Westheimer[20−22] was used for many years for carboxylic acids. The Kirkwood−Westheimer model treats the solute as elliptical in shape. More recently, numerical solutions of the Poisson−Boltzmann (PB) equation that allow a detailed representation of molecular shape have been shown to yield good agreement with measured substituent effects in dicarboxylic acids and diamines.[23,24] These results suggest that a purely electrostatic model is adequate to predict the polar substituent effects upon pK_as in these cases.

On the other hand, previous calculations of the pK_as of small molecules have rarely accounted explicitly for the conformational flexibility of the solutes. It is therefore of interest to examine whether the electrostatic model works when flexibility is accounted for. A recent molecular dynamics study with explicit solvent does account for three conformations of succinic acid in calculations of the difference between the two apparent pK_as of this molecule.[25] Such calculations may offer insights into the influence of the granularity of the solvent upon the energetics of ionization. However, it may be difficult to apply such approaches to more complex molecules, especially when their predominant conformations are not obvious *a priori*.

The present study accounts for solute flexibility by the MM method, which efficiently locates stable conformations and sums their configuration integrals to yield the total free energy (or, more properly, the chemical potential) of the molecule. These calculations are tractable largely because solvent is modeled with the rapid GB electrostatics approximation. This is, to our knowledge, the first application of either the GB model or the MM method to the computation of pK_a shifts. The molecules examined are dicarboxylic acids and aminocarboxylic acids.

pK_a Shifts in the Active Site of HIV-1 Protease. The Human Immunodeficiency Virus 1 protease (HIVP) possesses two aspartyl residues that form a chemically symmetric catalytic dyad.[26] The carboxyl groups of these two residues occupy a complex environment: the active site is partly sequestered from solvent, and the two carboxyl groups approach each other to within <3 Å (see, e.g., refs 26 and 27). The pK_as of the aspartyl dyad are known approximately from kinetic studies as a function of pH. It has recently been shown that numerical solutions of the PB equation can be used to compute the pK_as of the dyad with good accuracy.[28] Because the aspartyl dyad does not interact strongly with other ionizable groups in the protein, it can be viewed as a particularly complicated dicarboxylic acid. Its pK_as can thus be computed in the same way as the pK_as of the small dicarboxylic acids discussed above. Here, the MM/GB method is used to compute the pK_as of the aspartyl dyad in HIVP with all other ionizable groups artificially set neutral. The results are compared with experiment and with detailed solutions of the PB equation.

This paper is organized as follows. The Methods section explains the theory used to compute pK_as with the MM/GB method. This material is presented in some detail because it is general enough to serve as the basis for other methods. The computational methods are then described. The Results section compares computed results with experimental data and related calculations. The Discussion section considers the accuracy and the potential utility of the new methods studied here and analyzes the physical basis for polar substituent effects on pK_as.

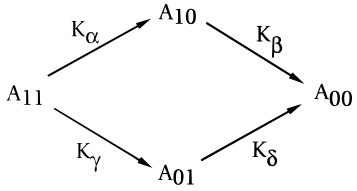

Figure 1. Protonation equilibria of an asymmetric, difunctional acid.

Methods

1. Theory. This subsection derives formulas for the apparent pK_as of flexible molecules bearing two ionizable groups. Classical statistical thermodynamics is assumed to be applicable; nonclassical contributions to the work of deprotonation are estimated by reference to model compounds.[17,20−25] An approach to formulating a hybrid quantum-classical model is also sketched.

Expressing Macroscopic pK_as in Terms of Configuration Integrals. The macroscopic pK_as of diacid A are given by $-\log K_1$ and $-\log K_2$, where K_1 and K_2 are the apparent acid dissociation constants for the following reactions:

$$AH_2 \overset{K_1}{\rightleftharpoons} AH + H^+$$

$$K_1 = \left(\frac{[AH][H^+]}{[AH_2]}\right)_{eq} \tag{1}$$

$$AH \overset{K_2}{\rightleftharpoons} A + H^+$$

$$K_2 = \left(\frac{[A][H^+]}{[AH]}\right)_{eq} \tag{2}$$

Here the "eq" subscript indicates a ratio of concentrations at equilibrium, and the concentrations are in units of "standard concentrations".[30] It is assumed that the activity coefficients of the reactants are unity.

The macroscopic reactions in eqs 1 and 2 obscure the fact that an asymmetric compound with two titratable groups possesses four different protonation forms. These will be referred to as A_{11}, A_{10}, A_{01}, and A_{00}, where subscripts of "1" and "0" indicate, respectively, the presence or absence of a proton on the two titratable groups. (Symmetric compounds, which require a slightly different treatment, are considered at the end of this subsection.) To relate the macroscopic equilibrium constants, K_1 and K_2, to microscopic properties, it is necessary to rewrite them in terms of the microscopic equilibrium constants of the molecule, as diagrammed in Figure 1, where K_α, K_β, K_γ, and K_δ are the equilibrium constants for the

(19) Smallwood, H. M. *J. Am. Chem. Soc.* **1932**, *54*, 3042.
(20) Kirkwood, J. G.; Westheimer, F. H. *J. Chem. Phys.* **1938**, *6*, 506.
(21) Kirkwood, J. G.; Westheimer, F. H. *J. Chem. Phys.* **1938**, *6*, 513.
(22) Kirkwood, J. G.; Westheimer, F. H. *J. Chem. Phys.* **1938**, *7*, 437.
(23) Potter, M. J.; Gilson, M. K.; McCammon, J. A. *J. Am. Chem. Soc.* **1994**, *116*, 10298−10299.
(24) Rajasekaran, E.; Jayaram, B.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 8238−8240.
(25) Figueirido, F.; Del Bueno, G.; Levy, R. M. *J. Phys. Chem.* **1996**, *100*, 6389−6392.
(26) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. *Science* **1989**, *245*, 616−621.
(27) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391.
(28) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. Submitted for publication.
(29) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.
(30) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.

indicated acid dissociations. Because [AH] = [A_{10}] + [A_{01}], it follows that

$$K_1 = K_\alpha + K_\gamma$$

$$K_2 = (K_\beta^{-1} + K_\delta^{-1})^{-1} \tag{3}$$

The microscopic equilibrium constants described above are directly related to the standard chemical potentials of the reactants and products.[1,31] For example,

$$K_\alpha = \exp[-\beta \Delta G_\alpha^\circ] \tag{4}$$

where

$$\Delta G_\alpha^\circ \equiv \mu_{10}^\circ + \mu_H^\circ - \mu_{11}^\circ \tag{5}$$

$$\beta \equiv (kT)^{-1}$$

Here ΔG_α° is the standard free energy of the acid dissociation, k is Boltzmann's constant, and μ_{10}°, μ_H°, and μ_{11}° are the standard chemical potentials of A_{10}, H^+, and A_{11}, respectively. Analogous equations hold for the other three microscopic equilibrium constants, K_β, K_γ, and K_δ. For molecular transformations that are described well by classical statistical thermodynamics, the relevant contributions to the chemical potential may be written in terms of an integral over molecular conformations, i.e., a configuration integral Z. For the purposes of this paper, a convenient form is[30]

$$\mu^\circ = -RT \ln\left(\frac{8\pi^2}{C^\circ \sigma} Z\right) \tag{6}$$

$$Z \equiv \int e^{-\beta(U(\mathbf{r})+W(\mathbf{r}))}\, d\mathbf{r} \tag{7}$$

Here C° is the standard concentration; σ is the symmetry number of the molecule; $U(\mathbf{r})$ and $W(\mathbf{r})$ are, respectively the gas-phase potential energy and solvation energy of the molecule as a function of its conformation; and $\mathbf{r}$ is a vector of internal coordinates that specify the conformation. (This formula neglects several contributions that will cancel in the present application.) From eqs 5−7, ΔG_α° can be written as

$$\Delta G_\alpha^\circ = -\mathrm{RT} \ln \frac{\sigma_{11}}{\sigma_{01}} \frac{\int e^{-\beta(U_{01}(\mathbf{r})+W_{01}(\mathbf{r}))}\, d\mathbf{r}}{\int e^{-\beta(U_{11}(\mathbf{r})+W_{11}(\mathbf{r}))}\, d\mathbf{r}} + \mu_H^\circ \tag{8}$$

Evaluating this formula is difficult, because it involves a change in covalent bonding—deprotonation of the acid—and also requires integrating over conformations. Empirical force fields and solvation models are helpful in evaluating the configuration integrals, but they cannot provide the energy changes associated with deprotonation. On the other hand, electronic structure calculations can yield energy changes associated with deprotonation, but they are too slow to be used for the integrals. It is therefore of interest to consider evaluating ΔG_α° with a hybrid quantum-classical approach. This approach is not used in the present paper, but it provides a theoretical basis for the empirical method that actually is used.

(31) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986.

A Hybrid Quantum-Classical Approach to Evaluating pK_as. This approach assumes that the potential energy as a function of conformation that is provided by an empirical force field equals, to within an additive constant, the true potential energy as a function of conformation. That is, it is assumed that the empirical force field correctly reproduces the dependence of energy upon conformation, even though it fails to reproduce the change in energy associated with deprotonation. (Existing force fields are, of course, only approximations to the true conformational energy surface.[32] On the other hand, the success of empirical force fields in a range of applications suggests that they are sufficiently accurate in energetically important regions of configuration space to be useful in the present application.) Thus,

$$U^f_{01}(\mathbf{r}) = U_{01}(\mathbf{r}) + C_{01}$$

$$U^f_{11}(\mathbf{r}) = U_{11}(\mathbf{r}) + C_{11} \tag{9}$$

where the superscript f signifies the energy computed with an empirical force field. It is also assumed that the solvation energy $W(\mathbf{r})$ can be approximated adequately by an empirical model, $W^f(\mathbf{r})$, such as the PB/γ[33] or GB/SA model.[8] With these assumptions it is readily shown that

$$\Delta G_\alpha^\circ = -RT \ln \frac{\sigma_{11}}{\sigma_{01}} \frac{\int e^{-\beta(U^f_{01}(\mathbf{r})+W^f_{01}(\mathbf{r}))}\, d\mathbf{r}}{\int e^{-\beta(U^f_{11}(\mathbf{r})+W^f_{11}(\mathbf{r}))}\, d\mathbf{r}} + \mu_H^\circ + C_{01} - C_{11} \tag{10}$$

The quantity $C_{01} - C_{11}$ corrects for the inability of the empirical force field to yield energy changes that result from deprotonation. This quantity can, at least in principle, be computed by subtracting the force-field energy from the energy obtained by electronic structure calculations for the same conformation of the deprotonated and protonated forms of the molecule. In addition, the standard chemical potential of the aqueous proton can be computed from its gas-phase partition function, plus the standard work of transfer from vacuum to solvent.[30] The configuration integrals in eq 10, which now involve only empirical force fields, can be computed by the MM method.[14] Thus, all the terms in eq 10 can be computed or found in the literature. How well such an approach would work in actual practice remains to be seen. The approach used in the present paper avoids electronic structure calculations and relies instead upon the experimentally measured pK_as of model compounds.

Model-Compound Approach to Computing pK_as. The present approach is an elaboration of the electrostatic model that was apparently introduced by Bjerrum[17] and that has been used in many subsequent studies. This section describes the electrostatic model and provides a theoretical basis for its use. Consider a monofunctional model compound whose ionization free energy ΔG_{ex}° is known. This experimental free energy can, as above, be written in terms of chemical potentials:

$$\Delta G_{ex}^\circ \equiv \mu_0^\circ + \mu_H^\circ - \mu_1^\circ \tag{11}$$

Here the subscript of μ° is 1 for the protonated form of the compound and 0 for the deprotonated form. The experimental

(32) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.

(33) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

free energy is now written in terms of classical configuration integrals computed with an empirical force field, along with a correction term:

$$\Delta G^\circ_{ex} = -RT \ln \frac{\sigma_1}{\sigma_0} \frac{\int e^{-\beta(U^{f_0}(\mathbf{r})+W^{f_0}(\mathbf{r}))}\, d\mathbf{r}}{\int e^{-\beta(U^{f_1}(\mathbf{r})+W^{f_1}(\mathbf{r}))}\, d\mathbf{r}} + \Delta G_{corr} \quad (12)$$

This equation will serve as the definition of the correction term. The key approximation of the model-compound approach, then, is that *the correction, ΔG_{corr}, is the same for all functional groups of a given type.* For example, the correction is assumed to be the same for all carboxylic acids linked to aliphatic chains, no matter what other substituents are present. This approximation is justified by the analysis in the previous paragraph, which shows that the correction term is given by

$$\Delta G_{corr} \approx C_0 - C_1 + \mu^\circ_H \quad (13)$$

It is indeed plausible that the quantities C_0 and C_1 will be similar for various instances of the carboxylic group. Clearly, though, the model-compound approximation must be applied judiciously. It will be accurate only for difunctional compounds where the influence of the second functional group upon the first group is accurately reproduced via the empirical force field, i.e., where through-bond induction is not important. The range of validity of this approximation is considered further in the Discussion section.

Given this approximation, the value of ΔG_{corr} obtained from a model compound may be used as a correction for a difunctional compound with a similar group. Thus, once ΔG_{corr} has been determined for an appropriate model compound via eq 12, it can be used to correct the free energy for a difunctional compound:

$$\Delta G^\circ_\alpha = -RT \ln \frac{\sigma_{11}}{\sigma_{01}} \frac{\int e^{-\beta(U^{f_{01}}(\mathbf{r})+W^{f_{01}}(\mathbf{r}))}\, d\mathbf{r}}{\int e^{-\beta(U^{f_{11}}(\mathbf{r})+W^{f_{11}}(\mathbf{r}))}\, d\mathbf{r}} + \Delta G_{corr} \quad (14)$$

The free energies of the other reactions in Figure 1 can be computed by similar formulas. The configuration integrals in these formulas can be computed with the MM method. The macroscopic equilibrium constants K_1 and K_2 are then computed with eq 3. These macroscopic equilibrium constants may be compared directly with experimental data.

p*K*$_a$s of Symmetric Molecules. The treatment above applies to an asymmetric molecule. An adjustment is required when the two functional groups are chemically equivalent to each other, i.e., when the whole molecule is rotationally symmetric. In this case there is only one type of monoprotonated species. This case may be treated in two different ways. Perhaps most intuitively appealing is to use the reaction scheme in Figure 1 and to write, from eq 3, that

$$K_1 = 2K_\alpha = 2K_\gamma$$
$$K_2 = K_\beta/2 = K_\delta/2 \quad (15)$$

However, it is more rigorous to recognize that the two monoprotonated species of a symmetric, difunctional compound are actually identical to each other. Therefore, the correct

$$A_{11} \xrightarrow{K_\alpha} A_{10/01} \xrightarrow{K_\beta} A_{00}$$

Figure 2. Protonation equilibria of a symmetric, difunctional acid. The sole distinct form of the monoprotonated molecule is termed $A_{10/01}$.

reaction scheme is that shown in Figure 2 and

$$K_1 = K_\alpha$$
$$K_2 = K_\beta \quad (16)$$

However, this view of the problem does not alter the final values computed for K_1 and K_2. This is because the symmetry number of A increases by a factor of 2 on going from A_{11} to $A_{10/01}$, and then falls by a factor of 2 on going from $A_{10/01}$ to A_{00}. Therefore, correct application of eq 16 yields the same numerical values of K_1 and K_2 as eq 15.

2. Computational Methods. This subsection details the computational methods used in the present study. The first part describes how chemical potentials are computed with the MM method. The second part describes the potential energy and solvation energy models. The final part describes how the various molecules are represented.

2.1. Computation of Standard Chemical Potentials. Chemical potentials are computed in terms of classical configuration integrals (eqs 6 and 7), by the previously described MM method.[14] This method is implemented in a local version of the program UHBD.[6] Briefly, the MM method uses the fact that the largest contributions to the configuration integral are from regions of configuration space in or near energy minima. A complete configuration integral Z is approximated as a sum of the contributions z_j from a finite number M of wells in the energy surface:

$$Z \approx \sum_{j=1}^{M} z_j \quad (17)$$

The contribution of energy-well j is computed via a simple Monte Carlo procedure[34] as

$$z_j \approx \frac{v_j}{N_j} \sum_{i=1} e^{-\beta(U_i+W_i)} \quad (18)$$

where v_j is the volume of a hyperrectangle encompassing energy well j; N_j is the number of random samples taken within the hyperrectangle; and $U_i + W_i$ is the energy of sample i within the hyperrectangle.

It is reasonable to assume that the conformational distributions of the "hard" bond and angle degrees of freedom of the molecules are essentially fixed in the problems addressed here. Under these circumstances, it is appropriate to treat bonds and angles as rigid and to sample over only dihedral degrees of freedom.[14,30] This simplification is used in the present calculations. The configuration integrals in the present study extend over the dihedrals of all rotatable bonds, except as otherwise noted.

A calculation is said to be converged when the free energy does not change by more than 1 ppm when the contribution of a new energy minimum is added. The dimensions of each energy well are defined by an energy cutoff of 167 kJ/mol above the base of the well. As previously described, a Metropolis Monte Carlo calculation can be used to estimate the contribu-

(34) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes. The art of scientific computing;* Cambridge University Press: Cambridge, 1989.

Table 1. Model Compounds for PKa Calculations

	Group 1		Group 2	
difunctional compd	model compd	pK_a	model compd	pK_a
pimelic acid	hexanoic acid	4.86		
adipic acid	pentanoic acid	4.84		
glutaric acid	butanoic acid	4.82		
6-aminohexanoic acid	hexanoic acid	4.86	pentylamine	10.60
5-aminopentanoic acid	pentanoic acid	4.84	butylamine	10.64

tions to the configuration integral of conformations not in the identified energy wells.[14] However, this yields a negligible correction for the molecules examined here, so this extra step is not used in the present study.

2.2. Energy Model. As noted above, the energy in a configuration integral can be separated into a potential energy, $U(\boldsymbol{r})$, and a solvation energy, $W(\boldsymbol{r})$.[30] Here, the CHARMm 22 empirical force field with polar hydrogens only is used for the potential energy.[35] Parameters are assigned with the program Quanta[36] and, where necessary, the ChemNote module of Quanta. A molecular dielectric constant of 2 is used to account for electronic polarizability,[37] as is customary in modeling polar substituent effects.

The Generalized Born (GB) model[8] is used for W. Two analytic versions of the GB model have recently been described;[10,38] the one used here is that of Qiu et al. When equal atomic coordinates, radii, and charges are used, the present implementation of the GB model agrees well with the published implementation[38] in the program Batchmin5.5:[39] the deviations are less than 0.004 kJ/mol for 30 different mono- and difunctional compounds. No solvation contribution related to molecular surface area is included in W. This is because the surface area of these molecules varies little with conformation. For example, the surface areas computed for the energy minima of pimelic acid range over 20 Å^2. This corresponds to a variation of only 0.4 kJ/mol when the commonly used surface tension of 0.021 (kJ/mol) Å^{-2} [33] is used. Except as otherwise noted, the GB solvation energies reported here are evaluated with the radius of each atom set to the mean of its CHARMm σ parameter and that of water (atom type OW, radius 1.575 Å), except that hydrogen radii are set to 1.2 Å. The solvent dielectric constant is set to 78.

2.3. Molecular Models. Difunctional Compounds. Table 1 lists the three diacids and two zwitterions studied here, along with the relevant model compounds. The list includes the longest difunctional compounds of each class for which all relevant model-compound data were found in the literature. It also includes the shortest compounds that could be treated by the present model. This limitation is a consequence of the fact that the CHARMm 22 force field scales 1−4 electrostatic interactions by a factor of 0.5. In contrast, the solvation model, which is based upon the Poisson equation, does not scale the change in solvation energy that results from the interaction of these charges. This leads to a physically unrealistic imbalance in the net energy associated with adding the new functional group. For example, if adding the new group creates a 1−4 charge−charge repulsion, the associated Coulombic term will be scaled, resulting in only a modest destabilization. However, the corresponding solvent-mediated interaction, which is stabilizing, is not scaled. The net effect of the scaled destabilizing Coulomb term and the unscaled stabilizing solvent term will be to stabilize the molecule, even though a repulsive interaction has been added. This imbalance is physically unrealistic on theoretical grounds, and it also leads to marked discrepancies relative to experiment. As a consequence, numerically reasonable results are not obtained when an added functional group places partial charges in a 1−4 relationship with a preexisting charge.

Table 2. Symmetry Numbers for pK_a Calculations

species	σ	σ_{ext}	σ_{int}
neutral diacid	2	2	1
singly ionized diacid	2	1	2
doubly ionized diacid	8	2	4
neutral aminocarboxylic acid	1	1	1
aminocarboxylic acid (−)	2	1	2
aminocarboxylic acid (+)	3	1	3
zwitterionic aminocarboxylic acid	6	1	6
neutral model acid	3	1	3
ionized model acid	6	1	6
neutral model amine	3	1	3
ionized model amine	9	1	9

The difunctional compounds are built with Quanta[36] in all-trans conformations. The resulting structures are relaxed with 200 steps of conjugate gradient minimization with only covalent bond terms in CHARMm 22. The corresponding model compounds are built by deleting the appropriate groups in the difunctional compounds and then reassigning atomic parameters. The identities and the experimental pK_as of the model compounds are listed in Table 1.

As noted in the Theory section, it is necessary to correct configuration integrals for symmetry. Accordingly, Table 2 lists the internal (σ_{int}), external (σ_{ext}), and overall symmetry numbers ($\sigma = \sigma_{int}\sigma_{ext}$) of the various molecular species. Internal symmetries are those which result from rotation of a symmetric part of the molecule. External symmetry numbers result from symmetry under rotation of the molecule as a whole. The symmetry numbers are based upon the following considerations. The model compounds have no external symmetry, but their terminal methyl group has 3-fold internal symmetry. The diacids possess 2-fold external symmetry in the doubly ionized and doubly neutral states. An ionized carboxylate group contributes a 2-fold internal symmetry when it appears in either a model compound or a diacid. The zwitterions possess no external symmetry in all states. An ionized amine group contributes 3-fold internal symmetry.

HIV Protease. The calculations are based upon a crystal structure, Protein Data Bank[40] entry 1HHP.[27] For simplicity, all ionizable groups other than the aspartyl dyad are fixed in their un-ionized states. However, comparison of PB calculations with this approximation and full PB titration calculations[28] show little effect upon the pK_as of the dyad. Hydrogen coordinates are assigned with the HBUILD command in CHARMm[41] and energy minimized by steepest descent up to 500 steps with only hydrogens allowed to move. To limit CPU usage during the MM calculations, a residue-based cutoff of 7.5 Å is used to obtain a truncated protein centered on the two aspartates. In order that the present results be comparable to previous

(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(36) Molecular Simulations Inc., Waltham, MA.
(37) Gilson, M. K.; Honig, B. H. *Biopolymers* **1986**, *25*, 2097−2119.
(38) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.* **1997**, *101*, 3005−3014.
(39) Mahamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
(40) Bernstein, F. C.; Koetzle, T. F.; Williams, T. F.; Meyer, G. J. B., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.
(41) Brunger, A. T.; Karplus, M. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 148−156.

Table 3. Computed and Experimental pK_{a_1} and pK_{a_2} of Diacids

	calcd		exptl	
compd	pK_{a_1}	pK_{a_2}	pK_{a_1}	pK_{a_2}
pimelic acid	4.53	5.49	4.49	5.43
adipic acid	4.51	5.62	4.42	5.42
glutaric acid	4.41	5.75	4.34	5.43

Table 4. Computed and Measured pK_{a_1} and pK_{a_2} of Aminocarboxylic Acids[a]

	calc		expt	
compd	pK_{a_1}	pK_{a_2}	pK_{a_1}	pK_{a_2}
6-aminohexanoic acid	4.36	10.78	4.37	10.80[b]
5-aminopentanoic acid	4.19	11.19	4.26	10.77[b]
4-aminobutanoic acid			4.03	10.56[b]
3-aminopropanoic acid			3.55	10.30[b]
2-aminoacetic acid				9.86[c]

[a] pK_{a_1} and pK_{a_2} are the pK_as of the carboxylic and amino groups, respectively. [b] Experimental data extrapolated to 0 ionic strength.[52] [c] Experimental data at 0.5 M ionic strength. Other amines show pK_a shifts of about +0.2 when ionic strength increases from 0 to 0.5 M.[52]

calculations of protein pK_as,[4,42−45] the molecular interior is assigned a dielectric constant of 4, and the model compound pK_a is set to 4.0. The use of a dielectric constant 4 for proteins implicitly accounts for the mobility of protein residues.[37,46,47] An artificial model compound is constructed from the side chain plus the main chain atoms of the aspartyl residues. Torsional sampling with the MM method extends over only the four side-chain torsions of the aspartyl groups.

Results

1. Small Difunctional Molecules. pK_a Calculations. Table 3 compares the calculated and measured values of pK_{a_1} and pK_{a_2} for the three diacids. All results correspond to zero ionic strength. The agreement between calculation and experiment is good. Thus, the computed values of pK_{a_1} are similar to the experimental values and show the same trend with chain length. The computed values of pK_{a_2} are close to but a bit higher than experiment, especially for the shorter compounds. They decline with increasing chain length, whereas the measured pK_{a_2} values are fixed.

Table 4 compares computed and measured pK_as for the aminocarboxylic acids. The pK_{a_1} values, which are associated with the carboxylic groups, are found to be somewhat more acidic than the corresponding model compounds, in agreement with experiment. This is as expected, because the carboxylic acids titrate in the presence of a cationic ammonium group, which presumably provides electrostatic stabilization to the ionized form of the acid. The decrement of the computed carboxylic pK_as with chain length is correct in sign but slightly exaggerated. The same pattern was observed for the dicarboxylic acids (see above).

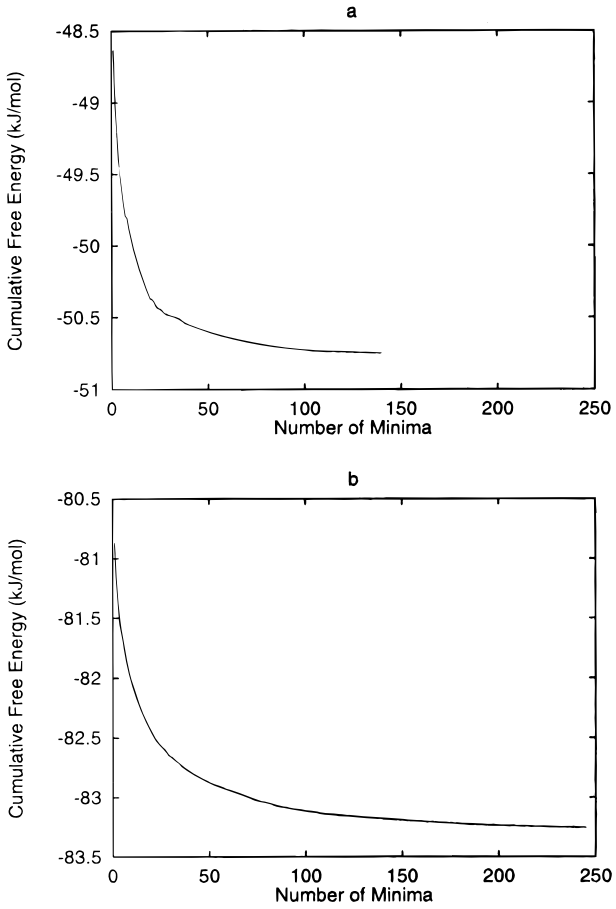

Figure 3. Convergence of free energy calculations for singly ionized pimelic acid (a) and doubly ionized 6-aminohexanoic acid (b).

The computed pK_as of the amino groups are somewhat less accurate. These groups are correctly predicted to be less basic than the corresponding model compounds, but they are too basic relative to experiment. Moreover, the computed trend of the pK_as with chain length is opposite to that seen experimentally. Experimentally, the molecule with the shorter aliphatic chain has the more basic amino group. This is unexpected, because the shorter chain presumably places the amino group closer to the anionic carboxylate. Experiments show that shorter molecules in the same series continue this unexpected trend as shown in Table 4. This observation is analyzed in the Discussion section.

Conformational Distributions. The present calculations yield not only pK_as but also the distribution of conformations. We examine these distributions by calculating the Boltzmann-weighted mean end-to-end distances for a series of singly ionized and doubly ionized diacids and for the zwitterionic forms of the amino acids. The results are plotted as a function of chain length in Figure 4. The zwitterions are consistently the most compact and the dianions the most elongated because of the attraction and repulsion, respectively, between the chain termini of these molecules. However, the length differences are small for the longer chains. This presumably results from the increasing dominance of chain entropy over electrostatics.

Although interactions between the polar termini do influence the mean chain lengths of these molecules, their influence is relatively weak. Thus, some extended forms of the zwitterions are nearly as stable as the salt-bridged forms. Also, it has been proposed that intramolecular hydrogen bonding influences the pK_as of dicarboxylic acids.[48] However, the present calculations yield no energy minima with such hydrogen bonds, even for

(42) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5804−5808.
(43) Yang, A.-S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins: Struct. Funct. Genet.* **1993**, *15*, 252−265.
(44) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415−436.
(45) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819−7833.
(46) Simonson, T.; Perahia, D.; Brunger, A. T. *Biophys. J.* **1991**, *59*, 670−690.
(47) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082−1086.

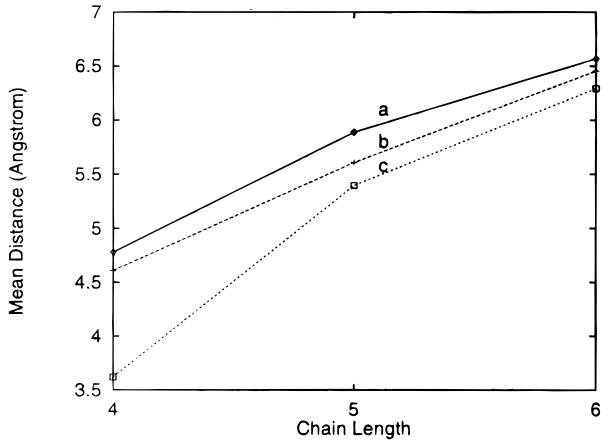

Figure 4. Mean distances for doubly ionized diacids (a), singly ionized diacids (b), and doubly ionized amino acids (c) versus the chain length in the number of bonds. The distances are between the carboxylic carbons for diacids, and between the carboxylic carbon and nitrogen for amino acids.

Table 5. Numbers of Energy-Minima (*M*) and CPU Times Required to Compute Converged Free Energies for Difunctional Compounds

compd	*M*	time (s)
pimelic acid	100	15100[a]
adipic acid	50	10800[a]
glutaric acid	35	6500[a]
6-aminohexanoic acid	224	18400[b]
5-aminopentanoic acid	101	7400[b]

[a] Computed with SGI R4400 Indigo 2. [b] Computed with SGI R10000 Indigo 2.

Table 6. Experimental and Computed pK_as of Aspartyl Dyad of HIVP with All Other Ionizable Groups Treated as Un-ionized[a]

	expt 1[b]	expt 2[c]	expt 3[d]	radii 1	radii 2
pK_{a_1}	3.4−3.7	3.3	3.1−3.3	5.44	4.09
pK_{a_2}	5.5−6.5	6.8	4.9−5.3	7.47	7.25

[a] Radii 1: default atomic radii defined as the mean of CHARMm 22 R_{min} values of atom and of water oxygen. Radii 2: radii defined as CHARMm 22 R_{min} values. [b] From the apparent catalytic constant.[53] [c] From the apparent catalytic constant.[54] [d] From the apparent inhibition constant.[53]

the neutral and monoanionic forms of these molecules. In summary, extended conformations are prominent for all of the difunctional compounds examined here. This result accounts for the success of previous pK_a calculations for the diacids that assumed fixed, extended conformations.[23,24]

Convergence and CPU Timings. The convergence behavior of the free energy calculations in the pK_a predictions is excellent. This is illustrated in Figure 3, which plots cumulative free energy as a function of the number of minima accumulated for the longest diacid and aminocarboxylic acid. Computer timings are listed in Table 5.

2. HIV Protease. pK_a Calculations. Table 6 compares aspartyl dyad pK_as computed with the MM/GB method with the results of three experimental studies. Calculations with the default atomic radii, marked "radii 1" in the table, markedly overestimate pK_{a_1}. However, the computed gap of about 3 between pK_{a_1} and pK_{a_2} agrees well with experiment. These results imply that the calculations overestimate the cost of ionizing each individual aspartyl group in the protein relative to solution, but that they accurately reproduce the repulsion between the two aspartyls. Both experiment and calculations show that this repulsion is stronger for the aspartyl dyad than for the small dicarboxylic acids (see above). This difference presumably results primarily from the closeness of the aspartyl groups in HIVP.

Table 7. pK_as Computed with MM/GB for Two Sets of Radii[a]

	radii 1		radii 2		exptl	
compd	pK_{a_1}	pK_{a_2}	pK_{a_1}	pK_{a_2}	pK_{a_1}	pK_{a_2}
pimelic acid	4.56	5.57	4.69	5.75	4.49	5.43
glutaric acid	4.41	5.75	4.49	6.10	4.34	5.43

[a] See text and Table 6 footnotes for details.

Table 8. Influence of the Cutoff in the Computation of pK_as of Aspartyl Dyad of HIVP[a]

	radii 1		radii 2	
	7.5 Å	no cutoff	7.5 Å	no cutoff
pK_{a_1}	5.23	5.32	3.81	2.41
pK_{a_2}	7.02	7.19	6.89	5.86

[a] See text and Table 6 footnotes for Details.

Because pK_as computed with radii 1 are not particularly accurate, the calculations were repeated with a slightly different set of atomic radii ("radii 2" in Table 6). These radii equal the R_{min} values from the CHARMm 22 parameters, except that polar hydrogens are still assigned a radius of 1.2 Å. With this set of radii, the carboxyl oxygens become somewhat smaller and carbons become somewhat larger, relative to radii 1. The new MM/GB results still overestimate pK_{a_1}, but they are close to the upper limit of the experimental data and yield a gap between pK_{a_1} and pK_{a_2} well within the experimental results. Thus, the agreement with experiment is better with the modified radii.

The results presented so far show that, although radii 1 worked well with MM/GB for the small diacids, they seem to overestimate the pK_as of the aspartyl dyad in HIVP. It is therefore of interest to determine whether radii 2, which work well with MM/GB in HIVP, provide good agreement with experiment for the diacids. We therefore repeated the MM/GB calculations for glutaric and pimelic acids with radii 2. The results, in Table 7, are fairly good, but they do not agree with experiment as well as the calculations with radii 1 (Table 3). In particular, the gap between pK_{a_1} and pK_{a_2} is overestimated even more than before.

It is worth examining the influence of the cutoff used in the HIVP calculations. This is done here by recomputing the pK_as of the aspartyl dyad for a fixed crystal conformation with and without the cutoff. These calculations are done for both radii 1 and radii 2, and the results are shown in Table 8. For radii 1, the cutoff has little effect upon the computed pK_as. However, for radii 2, pK_{a_1} rises by 1.4 when the cutoff is imposed, and pK_{a_2} rises by about 1.0. These results suggest that if the cutoff were removed in the full MM/GB calculations, the pK_as computed with radii 2 would fall well within the experimental range.

Conformational Distributions. The aspartyl groups of the dyad are packed in by other residues which are not treated as mobile in this study. As a consequence, the dyad possesses little conformational freedom: its energy minima are narrow and vary only by 180° rotations of the carboxyl groups. Moreover, because the ionized forms of the groups are symmetric to such rotations, they possess only one distinct energy minimum. Due to these conformational limits, convergence of free energy calculations is trivial to obtain.

(48) Dygert, S. L.; Muzii, G.; Saroff, H. A. *J. Phys. Chem.* **1994**, *74*, 2016−2026.

Discussion

This paper describes computations of the pK_as of small, difunctional molecules and of the aspartyl dyad in HIV protease. A novel method is used that accounts for the coupling between protonation and conformation, and the conformational distribution of the molecule is recomputed for each protonation state. This study also represents the first use of the GB model to compute pK_a shifts. The agreement with experiment is quite good in each case, supporting the basic validity of the approach. The following subsections discuss the small-molecule and HIV protease calculations in more detail.

1. Polar Substituent Effects in the Difunctional Molecules. Comparison with Previous Poisson−Boltzmann Calculations. Several studies have used the finite difference PB method to compute pK_a shifts for dicarboxylic acids.[23,24] The results have agreed well with experiment, despite the assumption of rigid, extended conformations. This is the first study to apply the GB model to these molecules, and the results are quite similar to those obtained with the PB model. However, because the GB model is more computationally efficient, it is readily combined with conformational analysis. This makes it possible to account for the coupling between protonation state and conformational distribution. We find that the conformational distribution of the dicarboxylic acids does not change markedly with protonation state. This accounts for the success of the previous PB calculations which assumed fixed conformations. Our calculations do yield significant occupancy of salt-bridged conformations for the zwitterionic forms of the amino acids. However, there are no previous finite difference PB calculations for these molecules.

The Physical Basis for Polar Substituent Effects. As noted in the Results section, the computed pK_as for the amine groups of the zwitterions are less accurate than those for the carboxyl groups. Moreover, the computed amino pK_a rises with decreasing chain length, whereas the experimental amino pK_a falls. Thus, the experimental results run counter to the expectation that the amino group will be more basic when it is closer to the carboxylate. This unexpected trend continues as the chain continues to shorten, as shown in Table 4, and as previously noted.[49] Computations were not done for these shorter chains for technical reasons (see Methods), but a similar electrostatic model does predict increasing basicity as the amine approaches the carboxylate.[49]

The trend cannot be adequately explained by increased salt-bridging in the longer compounds. This is because 4-aminobutanoic acid forms the most stable salt-bridges (see Figure 4), but the longer compounds are more basic. Also, salt-bridging is expected to make the amine more basic than a model compound lacking the carboxylate group; therefore, salt-bridging could not explain the observation that the shorter amino acids in this series are *less* basic than the model compounds, whose pK_as are about 10.6 (Table 1).

It seems more likely that the reduction in the pK_as of the amino groups in the short aminocarboxylic acids results from an additional, nonelectrostatic induction, such as the through-bond influences that are discussed in the literature on polar substituent effects (see, e.g., refs 15 and 16). Thus, as shown in Table 9, the pK_as of amino alcohols and of aminocarboxylate esters are lower than those of unsubstituted amines. Sample electrostatic calculations do not reproduce these shifts (results not shown). Perhaps, then, the pK_as of the amino groups in the aminocarboxylic acids are determined by a combination of nonelectrostatic and electrostatic influences, as previously suggested for glycine.[50]

Table 9. Measured pK_as of Hydroxylamines (0.5 mol/L ionic strength) and of Aminocarboxylate Esters (1 mol/L ionic strength)

compd	pK_a[a]	compd	pK_a[b]
5-aminopentan-1-ol	10.91		
4-aminobutan-1-ol	10.61		
3-aminopropan-1-ol	10.37	methyl β-alaninate	9.25
2-aminoethanol	9.80	ethyl glycinate	7.65

[a] Reference 55. [b] Reference 56.

Interestingly, in the piperidinemonocarboxylic acids these putative electrostatic and nonelectrostatic influences appear to cancel quite precisely.[51] As the carboxyl substituent is moved from position 2 to 4, the pK_a of the amine is essentially fixed. However, the pK_a of the carboxyl does change, as would be predicted by a through-space electrostatic model. This unexpected constancy of the pK_as of the amine groups was previously attributed to the formation of a hydrogen bond with the carboxyl when the two groups are close.[51] However, this explanation does not seem helpful: such a hydrogen bond would preferentially stabilize the cationic ammonium, and thus reinforce the expected electrostatic effect, rather than accounting for its absence. It seems more likely that the lack of variation in the amine pK_a over the three isomers results from the cancellation of electrostatic and nonelectrostatic influences of the carboxyl upon the amine.

If carboxylates make amines more acidic via nonelectrostatic induction, do they act in the same manner on other carboxylic groups? Interestingly, the calculations and experimental data for the dicarboxylic acids (Table 3) are indeed consistent with this hypothesis. This is because they show that the calculations, which consider only electrostatic influences, underestimate the acidity of the dicarboxylic acids. Perhaps, then the deviations of these electrostatic calculations from experiment result from the neglect of nonelectrostatic interactions between the groups, rather than from an inadequate treatment of electrostatics. The precise nature of any such nonelectrostatic interactions is uncertain. However, it is worth noting that, if they are not "through-bond" effects, they may be important in computing the pK_as of ionizable groups in proteins.

2. The Aspartyl Dyad of HIV Protease. The calculations for HIVP test the validity of the GB model in the complex environment of an enzyme active site. It is of some concern that a change of atomic radii is needed on going from the small molecules to HIV protease in order to retain accuracy. This observation suggests that the GB model needs adjustment if it is to yield highly reliable results for both small and large molecules. On the other hand, the overall quality of the results is quite good, especially given the simplicity of the model. This suggests that it may be possible to use the GB model to compute protein pK_as efficiently with a static model of the protein. The GB model may also allow protein pK_as to be computed with at least a partially flexible model of the protein.

(49) Edward, J. T.; Farrell, P. G.; Job, J. L.; Poh, B. *Can. J. Chem.* **1978**, *56*, 1122−1129.
(50) Saroff, H. A. *J. Chem. Educ.* **1994**, *71*, 637−643.
(51) Hansen, L. D.; Lewis, E. A. *J. Phys. Chem.* **1973**, *77*, 286−289.
(52) Martell, A. E.; Smith, R. M. *Critical Stability Consants*; Plenum Press: New York, 1974; Vol. 1-IV.
(53) Hyland, L. J.; Tomaszek, T. A., Jr.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454−8463.
(54) Ido, E.; Han, H.; Kezdy, F. J.; Tang, J. *J. Biol. Chem.* **1991**, *266*, 24349−24366.
(55) Alner, D. J.; Lansbury, R. C.; Smeeth, A. G. *J. Chem. Soc. A* **1968**, 417−421.
(56) Williams, A.; Jencks, W. P. *J. Chem. Soc. Perkin Trans. 2* **1974**, 1753−1759.

Acknowledgment. We are grateful to Dr. P. Shenkin for his patient help with implementing the GB method. This work was supported by the National Institute of Standards and Technology and the National Institutes of Health (GM54053 to M.K.G. and GM31034 to J.M.). Certain commercial equipment or materials are identified in this paper to specify the methods adequately. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

JA974307I