Small Molecule pK_a Prediction with Continuum Electrostatics Calculations

Michael J. Potter,*,† Michael K. Gilson,‡ and J. Andrew McCammon†

Department of Chemistry, University of Houston Houston, Texas 77204-5641
Center for Advanced Research in Biotechnology National Institute of Standards and Technology 9600 Gudelsky Drive, Rockville, Maryland 20850

Received August 18, 1994

In the simulation of biomolecules and their interactions, it is critically important to correctly model the protonation states of titratable groups. Ideally, such models would allow for adjustment of the protonation states to reflect conformational and other environmental changes. It is thus desirable to develop theoretical methods capable of quickly and accurately predicting pK_a values. A number of technique based on Poisson–Boltzmann electrostatics calculations have recently been described for application to proteins.[1] In the method of Antosiewicz et al.,[1g] each group is assigned an "initial" pK_a which represents its behavior when the group is isolated in solution. Finite-difference solutions to the linearized Poisson–Boltzmann equation are then used to adjust this pK_a to reflect the group's electrostatic environment within the protein. To test this approach further and to explore its utility for other types of molecules, we have applied the method to a variety of experimentally well-characterized diamines and diacids. The compounds considered were ethylenediamine, 1,3-diaminopropane, 1,2-diaminopropane, 1,4-diaminobutane, malonic acid, succinic acid, and glutaric acid.

The molecules were built in all-trans conformations using QUANTA 4.0 and energy-refined (200 steps of conjugate gradient minimization with CHARMm 22.0)[2] to relax stresses that might have resulted from the building process. The model compounds were created by replacing the appropriate groups in the difunctional compound structures. Only trans conformations were considered, except as noted below. This was done to test the limits of ignoring conformational flexibility and because it was expected that the all-trans conformation would be the dominant form in the doubly-ionized state. This assumption has been supported in previous theoretical electrostatic studies of small molecules.[3]

The electrostatic method used was similar to that of Antosiewicz et al.[1g] A "model" compound for each group was chosen on the basis of structural similarity. In general, the amine or carboxylic acid with the appropriate number of carbons was used, i.e., ethylamine for ethylenediamine, etc.[4] The molecules were represented as low dielectric regions ($\epsilon = 2$)[5] containing point charges at the centers of the atoms. Charges and radii (0.5σ) were taken from the OPLS parameter set.[6] Only polar hydrogens were included. The Richards probe-accessible surface definition was used to determine the extent of the low dielectric region.[7] Dielectric boundary "smoothing" was used.[8] The molecules were immersed in a high dielectric solvent ($\epsilon = 78$). Electrostatic potentials and energies were determined from finite-difference solutions to the linearized Poisson–Boltzmann equation calculated with the program UHBD.[9] Finite-difference focusing[7b] using two $65 \times 65 \times 65$ grids with spacings of 0.5 and 0.15 Å was employed. Ionization was modeled either as the addition of a positive or negative charge to the nitrogen of amino groups or the carboxyl carbon in the acids, respectively, or through the use of partial charges appropriate for the group and its ionization state.[10] For the ±1 ionization model, an "intrinsic" pK_a was calculated for each group using the following formula: $pK_{a_{int}} = pK_{a_{model}} - z[\Delta\Delta G/(2.303RT)]$. Here, z is ±1 for bases and acids, respectively, R is the gas constant, T is the temperature, and $\Delta\Delta G$ is the difference in electrostatic energy of ionization between the "model" compound and the site within the difunctional molecule with all other sites neutral:[1g]

$$\Delta\Delta G = {}^1/_2(\Psi_{11} - \Phi_{11}) + z\sum_{a=1}^{n} q_a\Psi_{1a} - z\sum_{a=1}^{m} q_a\Phi_{1a} \quad (1)$$

Here, atom 1 is the ionization site, Ψ_{1a} and Φ_{1a} are the potentials created at atom a by a unit positive charge on the ionization site in the difunctional compound and model, respectively, n and m are the numbers of atoms in the difunctional compound and the model, and q_a is the charge on atom a in the neutral form of the molecule. Model compound pK_a values were taken from Martell and Smith.[11,12] The difunctional compounds' pK_a values were computed by standard formulas[13] from $pK_{a_{int}}$ and the interaction energy between sites, $\Delta\Delta G_{inter}$, which is equal to the potential created at one ionization site by a unit positive charge at the other.

For the partial charge ionization model, the formulas used were as follow: $\Delta\Delta G = (G_1^d - G_0^d) - (G_1^m - G_0^m)$; $\Delta\Delta G_{inter} = (G_2^d - G_1^d) - (G_1^d - G_0^d)$. Here, G_2^d, G_1^d, and G_0^d are the electrostatic energies of the diprotonated, monoprotonated, and neutral forms of the difunctional compound, respectively, and G_1^m and G_0^m are the electrostatic energies of the protonated and neutral forms of the model compound.

In the case of ethylenediamine, the partial charge model was extended to include consideration of multiple conformations. Two gauche and one trans conformer were considered for each ionization state. Their electrostatic energies were Boltzmann

† University of Houston. After January 1, 1995: Department of Chemistry and Biochemistry and Department of Pharmacology, University of California at San Diego, La Jolla, CA 92093.

‡ Center for Advanced Research in Biotechnology.

(1) (a) Gilson, M. K.; Honig, B. H. *Proteins* **1988**, *3*, 32–52. (b) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *9*, 327–335. (c) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5804–5808. (d) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473–486. (e) Oberoi, H.; Allewell, N. M. *Biophys. J.* **1993**, *65*, 48–55. (f) Yang, A. S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins* **1993**, *15*, 252–265. (g) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.

(2) (a) Molecular Simulations Inc., Burlington, MA. (b) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1982**, *4*, 187–217.

(3) Edward, J. T.; Farrell, P. G.; Jon, J. L. *J. Chem. Phys.* **1972**, *57*, 5251–5256.

(4) The models used were ethylenediamine, ethylamine; 1,3-diaminopropane, propylamine; 1,4-diaminobutane, butylamine; 1,2-diaminopropane, propylamine and isopropylamine; malonic acid, acetic acid; succinic acid, propionic acid; glutaric acid, butyric acid.

(5) While Antosiewicz et al.[1g] found calculations using an interior dielectric of 20 to be most accurate in proteins, $\epsilon = 2$ is more appropriate in this case, as in the absence of intervening rotatable dipoles, only the electronic polarizability should be included.

(6) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(7) (a) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176. (b) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327–335.

(8) (a) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591–3600. (b) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 909–912.

(9) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187–197.

(10) For the partial charges model, the addition of a hydrogen to or loss from the ionization site was modeled explicitly, i.e., the structures used contained the number of hydrogens corresponding to the ionization state. For the ±1 model, the un-ionized structure was used for all calculations.

(11) Martell, A. E.; Smith, R. M. *Critical Stability Constants*; Plenum Press: New York, 1974; Vols. I–VI.

(12) Model pK_a values were ethylamine, 10.636; propylamine, 10.566; butylamine, 10.643; isopropylamine, 10.67; acetic acid, 4.757, propionic acid, 4.874, butyric acid, 4.819.

(13) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W. H. Freeman and Company: San Francisco, CA, 1980; Vol. III, pp 850–852.

0002-7863/94/1516-10298$04.50/0

Table 1. Calculated pK_a Values[a]

compound	±1 model		partial charge		exptl[b]	
	pK_{a_1}	pK_{a_2}	pK_{a_1}	pK_{a_2}	pK_{a_1}	pK_{a_2}
ethylenediamine	7.16	10.21	6.81	9.90	6.848	9.928
ethylenediamine[c]			6.35	10.32	6.848	9.928
1,2-diaminopropane	6.71	10.16	3.98	9.62	6.61	9.72
1,3-diaminopropane	9.20	10.76	8.77	10.40	8.48	10.49
1,4-diaminobutane	9.80	10.91	9.81	10.55	9.20	10.65
malonic acid	−0.06	10.21	2.96	2.92	2.847	5.896
succinic acid	3.32	6.81	3.93	6.37	4.207	5.636
glutaric acid	4.29	5.94	4.54	5.42	4.34	5.43

[a] Numerical errors are estimated by varying grid size and spacing to be less than 0.1 unit for the ±1 model and 0.3 unit for the partial charge model. [b] Experimental data from Martell and Smith as reported in ref 11. Errors, when reported, are less than 0.1 unit in all cases. [c] Multiple conformations included in the calculation.

Table 2. Breakdown of Results[a]

compound	±1 model		partial charge		exptl[b]	
	$pK_{a_{int}}$	ΔpK_a	$pK_{a_{int}}$	ΔpK_a	$pK_{a_{int}}$[c]	ΔpK_a
ethylenediamine	9.91	3.05	9.60	3.09	9.627	3.08
ethylenediamine[d]			10.02	3.97	9.627	3.08
1,3-diaminopropane	10.46	1.56	10.10	1.63	10.19	2.01
1,4-diaminobutane	10.61	1.11	10.25	0.74	10.34	1.45
malonic acid	0.24	10.18	3.26	0.04	3.148	2.849
succinic acid	3.62	3.49	4.23	2.44	4.508	1.429
glutaric acid	4.59	1.65	4.84	0.88	4.641	1.090

[a] Numerical errors are estimated by varying grid size and spacing to be less than 0.1 unit for the ±1 model and 0.3 unit for the partial charge model. [b] Experimental data from Martell and Smith as reported in ref 11. Errors, when reported, are less than 0.1 unit in all cases. [c] Experimental pK_{a_2} (diamines) or pK_{a_1} (diacids) values corrected for the statistical factor. [d] Multiple conformations included in calculation.

averaged (based for simplicity on the continuum electrostatic free energy) for each state and subsequently used in the expressions above.

Tables 1 and 2 compare experimental pK_a values with those computed with the ±1 model. This is the model most similar to that used in the protein calculations. The results are surprisingly good for the diamines, with errors less than 0.8 unit in all cases. The diacid results are quite poor, though they improve as the distance between groups increases. Table 2 presents the results in such a way as to allow the independent evaluation of the calculation of $pK_{a_{int}}$ and the interaction energy (which determines $\Delta pK_a = pK_{a_2} - pK_{a_1}$). In the diamine calculations, the interaction energy is slightly underestimated. This is likely due to the crudeness of the ±1 ionization model and the inclusion of only the trans conformation. Consideration of other conformers should lead to an increased interaction energy. In contrast, the interaction energy is greatly overestimated in the diacids, probably due to the inadequacy of representing ionization of a carboxylic acid by a −1 charge on the carboxyl carbon. For the amines, the calculated $pK_{a_{int}}$ values are in reasonable accord with experiment. However, while the correct direction of change from the models is predicted for the diacids, the magnitude of the change is too large. Overall, while the results are reasonable for the diamines and at large separations (as in proteins) for the diacids, problems were encountered in the small diacids which are likely the result of the simple ionization model.

Results for the partial charge model are also given in Tables 1 and 2. They are slightly better for the amines and much better for the diacids. The reason appears to be that the intrinsic pK_a values are much more accurate than those obtained from the ±1 model. The interaction energies for the diamines are similar to those obtained with the first model. The interaction energy results for the diacids, though better than those for the ±1 model, are still not particularly good. Indeed, it is predicted that ionization in malonic acid is slightly cooperative! These problems likely reflect the inadequacy of a simple molecular mechanics model in the case where the two titrating groups are separated by only one methylene and indicate a limitation in the range of validity of the procedure. It is interesting that the addition of a single methylene on going from malonic to succinic acid produces so significant an improvement in the results.

Results for the multiple conformation treatment of ethylenediamine are given in Tables 1 and 2. This is the most detailed of the models considered. The predicted pK_a values are good but slightly worse than those obtained using the other two models. The interaction energy is slightly overestimated, and the $pK_{a_{int}}$ found is approximately 0.4 unit higher than the experimental value. While one would expect an elaboration of the model to result in better predictions, there are reasons why this may not be the case here. The multiconformation calculation involved 11 evaluations of the electrostatic energy, while the ±1 model required only two and the partial charges model five. The combination of numerical errors from the larger number of energy evaluations could worsen the results. It is also true that the treatment of conformational flexibility used here was very simple. No attempt was made to account in detail for ethylenediamine's ability to form intramolecular hydrogen bonds and the effect of solvent on these bonds or for solute electronic structural effects on the rotamer populations or interactions among the charge sites.

It appears that the purely electrostatic approach used in protein pK_a calculations is quite accurate, even in small molecules. In particular, excellent results were obtained for the intrinsic pK_a values of groups when the partial charge model was used. This is encouraging for protein calculations, as it indicates that the effect of a group's environment is accurately accounted for by the continuum electrostatic calculation. Though not included here due to space limitations, calculations were also performed for both the ±1 and partial charges models, in which $pK_{a_{int}}$ was set equal to $pK_{a_{model}}$. The results of these calculations were generally worse than those obtained as described above. This demonstrates that the results are indeed improved by the adjustment of $pK_{a_{model}}$ to account for the site's environment within the difunctional compound. Interaction energies are less well predicted. Conformational flexibility may play a small role in this, as evidenced by the underestimation of interaction energies in the amines. However, it appears that the majority of the problems are found in the diacids and are caused either by the simplicity of the ±1 ionization model or by the limitations of the molecular mechanics model. These concerns are less likely to be a problem in proteins as greater through-bond and direct intergroup distances are involved. The results also indicate that it is not unreasonable to attempt to extend these methods to the study of medium-sized organic molecules. It is clear that in such systems the ±1 model may not be accurate, especially for acids, and that care must be taken in cases where groups are close together.

Acknowledgment. This work has been supported in part by the National Institutes of Health, the National Science Foundation Metacenters Program, and the National Institute of Standards and Technology. M.J.P. is a Howard Hughes Medical Institute Predoctoral Fellow. M.K.G. was supported in part by a Howard Hughes Physician Research Fellowship.