Using a Charging Coordinate in Studies of Ionization Induced Partial Unfolding

Mitsunori Kato and Arieh Warshel*

Department of Chemistry, University of Southern California, 3620 McClintock Avenue, Los Angeles, California 90089-1062

Received: February 24, 2006; In Final Form: April 19, 2006

The ionization of groups in proteins may sometimes involve a partial unfolding and/or water penetration. Unfortunately the corresponding structural changes might not be captured by microscopic free energy perturbation (FEP) approaches due to activation barriers that are not surmounted in nanosecond FEP simulations. This problem is apparent, for example, from mutation experiments that introduced ionizable groups in protein interiors and from the difficulties to reproduce the corresponding pK_a changes by microscopic approaches. Here we develop a new general approach for treating such challenging cases. Our approach drives the protein structural change by increasing the charge of the ionized group beyond its physical value and thus overcoming the barriers for the partial unfolding by a physically consistent process. The potential of our approach is illustrated by the evaluation of the pK_a of the Val66Glu mutant of staphylococcal nuclease. In this case it is first demonstrated that standard FEP approaches give extremely disappointing results for this pK_a. On the other hand, our "overchargning" approach gives a much more realistic result. We believe that the present approach represents a breakthrough in FEP studies of ionizable residues in proteins, and expect this strategy to be useful in studies of a wide range of challenging problems including simulations of hydrogen exchange processes.

I. Introduction

Electrostatic energies play a major role in controlling protein function and in correlating the structure and function of biological molecules (e.g., refs 1−7). Yet the evaluation of the energetics of charged groups in proteins remains a significant challenge for microscopic models (e.g., refs 8−10) and the alternative use of macroscopic models is sometimes strongly dependent on the assumed dielectric constants.[11] The problem can be particularly serious when one deals with charges in protein interiors.[12]

One of the most dramatic demonstrations of the challenge in evaluation of the energetic of internal charges has been provided recently by the instructive works of Garcia-Moreno, Lattman, and their co-workers.[13−16] This research group provided fascinating examples of mutations of Val 66 to ionizable residues in the berried hydrophobic interior of staphylococcal nuclease (SNase). Surprisingly, and in contrast to most expectations, the pK_a of the mutants were not shifted in a drastic way as one would expect from charges in nonpolar environments. This finding appears to be inconsistent with regular macroscopic models,[15] and it indicates that the protein undergoes some local unfolding or water penetration. Of course, the real conceptual and practical challenge is to reproduce the observed pK_a shifts by some types of microscopic models without assuming a special dielectric constant.

Previous attempts[16] to use all-atom MD to study the effects of the mutations were quite instructive since they did not reproduce large water penetration in relatively long simulations (10 ns). This has indicated that there might be a barrier for the protein structural change, which is not surmounted during nanoseconds simulations. However, it has been hard to assess the full significance of the simulations since they did not involve attempts to evaluate the pK_a shift and thus left the problem unsolved.

To address the above challenge and the general problem of charge induced unfolding, we introduce a new method that uses the charge of the ionizable residue as an effective parameter that drives the unfolding process. This approach let us derive the protein structure along a reasonable partial unfolding coordinate and to explore charge-induced unfolding and water penetration processes. Our method appears to provide a new powerful tool for studies of diverse problems, ranging from the present pK_a problem to redox induced unfolding and hydrogen exchange.

II. Methods

Our challenge is to simulate the observed pK_a shift for the mutant of the Val166Glu mutation of SNase (or other ionizable residues in hydrophobic regions) by a microscopic all-atom approach. Here the starting point is the following relationship:[17]

$$pK_a^p = pK_a^w - \delta_0 \Delta\Delta G_{sol}^{w\rightarrow p}/(2.3\,RT) \qquad (1)$$

where ΔG_{sol} is the change in solvation energy upon moving from the un-ionized to the ionized form of the given group, $\Delta\Delta G_{sol}^{w\rightarrow p}$ is the change in ΔG_{sol} upon moving this group from solution to its site in the protein, and δ_0 is plus or minus for basic and acidic groups, respectively. Thus we have to reproduce $\Delta\Delta G_{sol}^{w\rightarrow p}$ by microscopic calculations.

The standard approach for such microscopic calculations is to use the free energy perturbation (FEP) in the form used in our original adiabatic charging approach.[8] This is done by using a mapping potential of the form

$$\epsilon_m(\lambda_m) = U_1(1 - \lambda_m) + U_2\lambda_m \qquad (3)$$

* Corresponding author. E-mail: warshel@usc.edu.

10.1021/jp061190o CCC: $33.50
Published on Web 05/20/2006

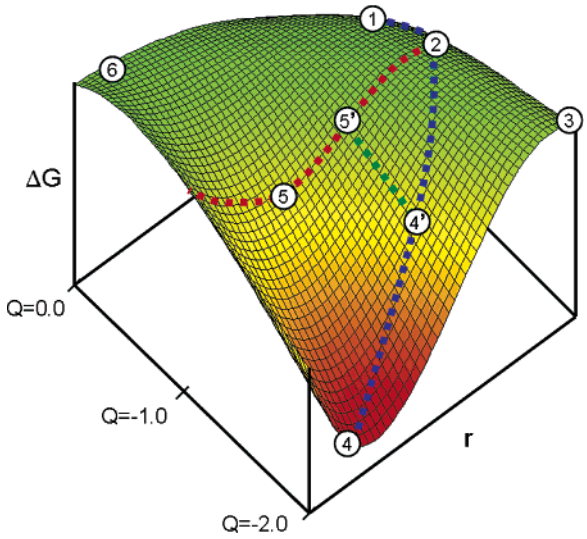

Figure 1. A schematic description of the overcharging procedure. The figure describes the free energy surface as a function of the charge of the ionizable residue and an effective protein coordinate, $\boldsymbol{r}$. Our mapping procedure is aimed at evaluating $\Delta G_{1\rightarrow 5}$, and this is accomplished by calculating $\Delta G_{1\rightarrow 5'} = \Delta G_{1\rightarrow 4'} + \Delta G_{4'\rightarrow 5'}$ and then finding the lowest $\Delta G_{1\rightarrow 5'}$.

where U_1 to U_2 are the potential surfaces of the system when the given ionizable group is neutral ($Q_1 = 0$) and fully ionized, ($Q_2 = Q_0$), respectively. λ_m is a parameter that changes between ($0 \leq \lambda_m \leq 1$). The free-energy increment, associated with the change of ϵ_m, can be obtained by [18]

$$\exp\{-\Delta G(\lambda_m \rightarrow \lambda_{m+1})\beta\} = \langle \exp\{-(\epsilon_{m+1} - \epsilon_m)\beta\}\rangle_m \quad (4)$$

where $\langle\rangle_m$ indicates that the given average is evaluated by propagating trajectories over ϵ_m. The overall free-energy change is now obtained by changing ϵ_m in n equal increments and evaluating the sum of the corresponding ΔG by

$$\Delta G(U_1 \rightarrow U_2) = \sum_{m=0}^{n-1} \Delta G(\lambda_m \rightarrow \lambda_{m+1}) \quad (5)$$

Here, however, we will use a special charging process to drive the system further from its regular final point $Q_2 = Q_0$ and continue the charging until we reach an unphysical large charges $Q_2 = Q_i$ that overcomes the unfolding barrier and reaches the landscape that will equilibrate eventually with the ionized form of the given group. Our concept is illustrated schematically in Figure 1 where we consider the effective protein coordinate ($\boldsymbol{r}$) and the charging coordinate (Q). This cycle leads:

$$\Delta G_{1\rightarrow 5} = \Delta G_{1\rightarrow 4} + \Delta G_{4\rightarrow 5} \quad (6)$$

Where $\Delta G_{1\rightarrow 5}$ is the correct free energy of the ionization process that includes the protein structural relaxation. The problem is that the points 5 and 4 might not correspond to a complete equilibrium condition and might undergo time-dependent relaxation. Thus we take the following approach: we first perform relatively short FEP simulations with different Q_i and collect the corresponding protein configurations $\boldsymbol{r}_i(Q_i)$. In the next step we perform longer FEP mapping but now with the constraint of being at $\boldsymbol{r}_i(Q = 0.0)$ at point 1, $\boldsymbol{r}_i(Q_i)$ at point 4′ and $\boldsymbol{r}_i(Q = -1.0)$ at point 5′. This leads to

$$\Delta G_{1\rightarrow 5} = \min\,(\Delta G_{1\rightarrow 4'} + \Delta G_{4'\rightarrow 5'}) \quad (7)$$

where "min" designates the lowest value. The relevant ΔGs are

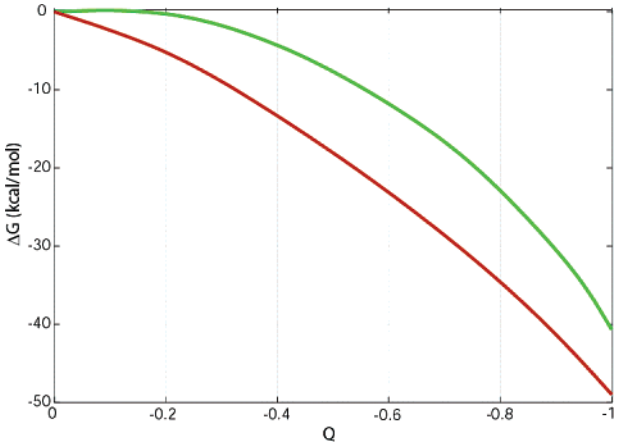

Figure 2. Showing the results of adiabatic charging FEP calculations for Glu 66 in SNase with the use of a nonpolarizable and polarizable force field. The green line corresponds to the use of a nonpolarizable force field while the orange line corresponds to the use of a polarizable force field.

evaluated by a FEP/umbrella sampling (see ref 19 for a similar approach) using

$$\epsilon'_m(\lambda_m) = (U_1(Q_1) + E_{\mathrm{con},1}(\boldsymbol{r}))(1 - \lambda_m) + (U_2(Q_2) + E_{\mathrm{con},2}(\boldsymbol{r}))\lambda_m \quad (8)$$

$$\text{where } E_{\mathrm{con},i}(\boldsymbol{r}) = \sum_k K(r_k - r^{\mathrm{o}}_{i,k})^2$$

and

$$\Delta g_i(z) = -\frac{1}{\beta}\ln[\exp\{-\Delta G(\lambda_m)\beta\}]\langle\delta(z - z_0)\exp\{-(U_i(z) - \epsilon_m(z))\beta\}\rangle_{\epsilon_m'} \quad (9)$$

Here $\Delta G(\lambda_m)$ is the result obtained by eq 5 for $\epsilon'_m \rightarrow \epsilon'_{m+1}$, z is a mapping coordinate defined by $z = |\boldsymbol{r} - \boldsymbol{r}_1{}^0|/|\boldsymbol{r}_2{}^0 - \mathrm{r}_1{}^0|$, $E_{\mathrm{con},m}(\boldsymbol{r}) = E_{\mathrm{con},1}(\boldsymbol{r})(1 - \lambda_m) + E_{\mathrm{con},2}(\boldsymbol{r})\lambda_m$, and $\Delta g_i(z)$ is the free energy functional for U_i. The charging free energy for $\Delta G(\boldsymbol{r}_1(Q_1) \rightarrow \boldsymbol{r}_2(Q_2))$ is given by the difference between the minimum of $\Delta g_2(z)$ and the minimum of $\Delta g_1(z)$.

The present study was performed on Staphylococcal Nuclease (PDB entry: 1U9R) using the MOLARIS simulation package.[20,21] The microscopic FEP calculations were performed using the ENZYMIX module of MOLARIS. The simulation included the use of 18 Å of the SCAAS spherical constraints[22,23] and the local reaction field (LRF) long-range treatment.[24] The FEP simulations were done with 31 frames that transformed the net charge of solute from 0.0 to its final value, where each frame included 50 ps of simulation with a 1fs time step.

III. Results and Discussion

To explore the performance of our method we examined the very challenging problem of the pK_a of Glu 66 in the hydrophobic site of SNase. As stated above the reproduction of the observed effect has been found at present a major challenge for macroscopic model. Furthermore, the attempt to perform long time simulations has not produced a significant water penetration, which means that there is a barrier for the protein structural change that is not surmounted by nanoseconds simulations.

In the first step in our study we attempted to use the regular FEP adiabatic charging procedure to evaluate $\Delta\Delta G_{\mathrm{sol}}^{\mathrm{w}\rightarrow\mathrm{p}}$ for Glu 66 (Figure 2). Not surprisingly, we obtained a major deviation

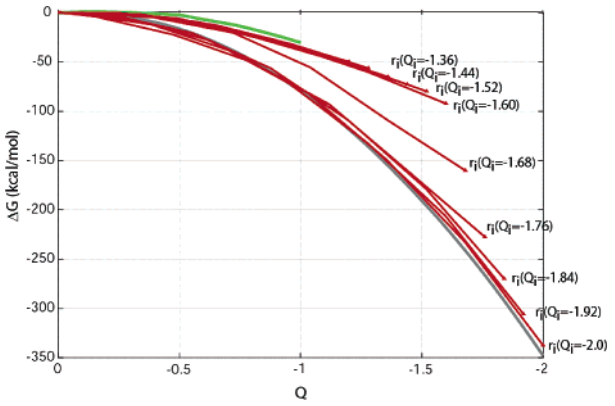

Figure 3. Illustrating the hysteresis for uncharging Glu 66 in the hydrophobic site of SNase from $Q = Q_i$ to $Q = 0$, starting from protein structures ($\boldsymbol{r}_i$s) generated by overcharging from $Q = 0$ to $Q = -2$. The green−blue lines are the result obtained for unconstrained forward charging from 0.0 to −1.0 and from 0.0 to −2.0, respectively. Note that out procedure for evaluating the pK_a of Glu 66 involves the use of the cycle of eq 7 (where as discussed in the text the final result is independet of the hysterisis shown here), rather than the procedure illustrated in this figure.

from the observed pK_a (predicted p$K_a \approx 23$ instead of about 9) despite running relatively long simulations of 5 ns. We also obtained a similar deviation while using polarizable force field (p$K_a \approx 19$). This indicates that the calculations could not reproduce the local unfolding and/or water penetration process, which are probably associated with the ionization of Glu 66. The problem is, of course, to capture this configurational change by the available simulation times.

The failure of the above FEP calculations indicted that we need a new approach that will induce the partial unfolding process, despite the fact that the system has a significant barrier for such a process. Here we felt that arbitrary attempts to unfold the protein may not be effective in reproducing the correct unfolding process and the correct physics of gradual water penetration. To address this challenge we used the overcharging approach described in the method section, where we considered the charging coordinate as a tool to induce a physically consistent unfolding process and as a way to surmount the activation barrier for the partial unfolding process. This approach allows one to extend the charging process to artificially large charges and thus overcome the unfolding barrier and reach the landscape that is populated by the ionized form of Glu 66, as demonstrate schematically in Figure 1.

The initial overcharging procedure and its effect is demonstrated in Figure 3 for the case of Glu 66 in SNase, for a nonpolarizable force field (the reason for using a nonpolarizable model will be discussed below). The figure describes a FEP procedure where we start from $Q = 0.0$ move to $Q = -1.0$ and continue to $Q = Q_i$ (where $|Q_i| > 1.0$). We then perform an uncharging procedure moving from $Q = Q_i$ to 0. (This procedure does not correspond to the procedure of eq 7). Performing such a charging and unchanging cycle leads to a remarkable hysteresis that can be seen by an inspection of Figure 3. The hysteresis appears to be time dependent, where a long simulation led to a further drop in the charging free energy. This effect can be understood by inspecting the schematic diagram of Figure 1 and realizing, for example, that stopping at point 4′ can lead, in longer simulation time, to a relaxation to point 4. This is, however, not a problem with our actual procedure that involves the use of constraints for the protein coordinate, $\boldsymbol{r}$. Note that it does not matter how we obtain the given $\boldsymbol{r}$ (for the given simulation length in Figure 3) since we are basically interested

TABLE 1: Free Energy Contribution to the Overcharging Cycle[a]

configuration	1	2	3	4	6
driving charge (Q_i)	−1.0	−1.6	−1.68	−1.68	−1.8
ΔG ($0 \rightarrow Q_i$)	−41	−98	−103	−203	−142
ΔG ($Q_i \rightarrow -1$)	0	54	61	152	104
$\Delta\Delta G_{cycle}$ ($0 \rightarrow -1$)	−41	−44	−42	−51	−38
$\Delta\Delta G_{cycle}$ ($0 \rightarrow -1$)′	−47	−50	−48	−57	−44
$\Delta\Delta G_{cycle}$ (AH $\rightarrow$ A$^-$)[b]	−44	−47	−45	−54	−141
p$K_{a,calc}$	19, 23[c]	16	18	11 (8.8)[d]	21

[a] Energies in kcal/mol. The charge (Q_i) used to "drive" the configurational changes is called here the "driving charge". The configurations 3 and 4 were generated for $Q_i = -1.68$ without and with induced dipoles, respectively. $\Delta\Delta G_{cycle}$ ($0 \rightarrow -1$) designates the free energy obtained by the cycle $1 \rightarrow 4' \rightarrow 5'$. $\Delta\Delta G_{cycle}$ ($0 \rightarrow -1$)′ includes an estimate of the effect of the induced dipoles. $\Delta\Delta G_{cycle}$ (AH $\rightarrow$ A$^-$) includes the effect of moving from a nonpolar neutral COO group to a polar COOH group. [b] The charging cycle in water give $\Delta\Delta G_{cycle}$ ($0 \rightarrow -1$)′ = −79 kcal/mol, $\Delta\Delta G$ (AH $\rightarrow$ A$^-$) = −70 kcal/mol. [c] This value was obtained with a nonpolarizable force field while all values reflect the effects of the protein induced dipoles. [d] (In bracket) the observed pK_a taken from ref 15.

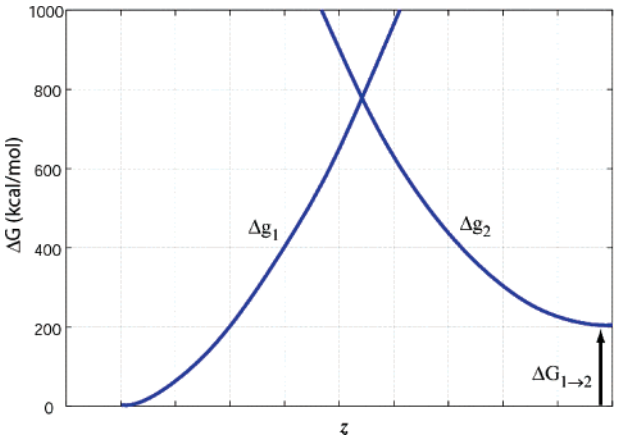

Figure 4. The free-energy functionals for the process of moving from $\boldsymbol{r}_1(Q = 0)$ to $\boldsymbol{r}_2(Q = -1.68)$. Δg_1 and Δg_2 are free energy functionals evaluated by eq 9.

in mapping along different values of $\boldsymbol{r}$ for Q = -1.0. At any rate it is important to note that we actually use the procedure of eqs 6−9, rather than the $Q = 0 \rightarrow Q = Q_i$ process described in Figure 3, which is only used to obtained the $\boldsymbol{r}_i$s.

Using the cycle defined in the method section (and eq 6) we evaluated the charging free energy for several values of Q_i with the results summarized in Table 1. The actual evaluation of the relevant Δg_i is illustrated in Figure 4 for $Q_i = -1.68$.

Before considering the calculated pK_as we have to clarify that in the case of charges in nonpolar sites, it is crucial to use polarizable force field,[1] and that this has been done in almost all our studies since 1976.[25] However in the present case, the situation is slightly more complicated than in usual studies with polarizable force fields (e.g., ref 12) since in contrast to the case without induced dipoles, the results in each direction depend on the charge at the final point. For example, charging up to $Q = -1.0$ with a final charge $Q_2 = -2.0$ gives different results than those obtained with $Q_2 = -1.0$; this is an inherent result of having U_2 with induced dipoles that are already polarized with Q_2. Therefore, our $\Delta g_i(Q_i,\boldsymbol{r}_i)$ will have different curvature with different values of the final charge even with the same $\boldsymbol{r}_i$. This problem may lead to ambiguity about the minimum of $\Delta g_i(Q_i,\boldsymbol{r}_i)$ and will probably require further studies. Thus, in the present case, we estimated the effect of induced dipoles, for each $\boldsymbol{r}_i$, by a regular FEP charging process, where

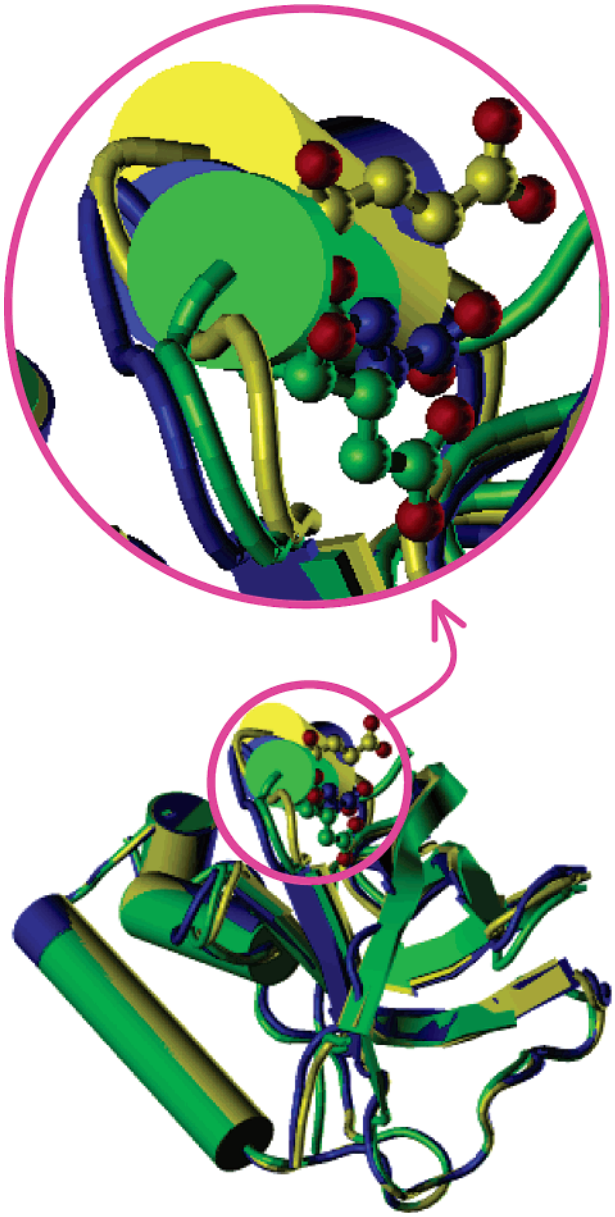

Figure 5. The structures obtained for different Q_is. Glu 66 is shown by balls and sticks in the cyan circle. The green structure is the native crystal structure at $Q_i = 0$, the yellow structure is the open structure obtained for $Q_i = -2$, and the blue structure is the half open structure obtained for $Q_i = -1.68$.

we evaluated $\Delta G(Q = 0 \rightarrow Q = -1.0, \boldsymbol{r} = \boldsymbol{r}_i)$. With this approximation, we obtained the fifth raw in Table 1 and the final pK_a values.

We also examined the structural changes induced by the overcharging (Figure 5) in order to be consistent with the experiment-based estimate that we must have only limited structural changes (e.g. see ref 16 and references in that work). As seen from Figure 5 the structures obtained with $Q_2 = -1.68$ represents a limited partial unfolding that seem to be consistent with the experimental analysis. (Although the actual nature of the structural changes has not been yet determined experimentally.) Thus we conclude that the results obtained with $Q_2 = -1.68$ give the best estimate of $\Delta G_{1\rightarrow 5}$. Combining all the above consideration, we conclude that the "correct" calculated pK_a corresponds to $\boldsymbol{r}_2$ ($Q_2 = -1.68$) and it is around 11 pK_a units.

Since the overcharging approach gave a large hysteresis for Glu 66 in SNase, it is justified to ask whether this approach will not change the results of pK_a calculations in more standard benchmarks or not. To explore this issue we repeated our approach for Glu 35 in Lysozyme. At this case, we did not obtain significant hysteresis, even after going to $Q = -2.0$ and using the resulting structure for the uncharging from $Q = -1.0$. We obtained p$K_a \approx 5$ in this calculation in a good agreement with the corresponding experimental results. Now although much more studies are needed for exploring our approach it seems that the overcharging approach allows one to sample the water penetration process and to see if it leads to a major change in the ionization process.

IV. Concluding Remarks

Microscopic evaluation of electrostatic energies in macromolecules can present a major challenge when the relevant charges are located in protein interiors. Typically microscopic calculations overestimate, drastically, the corresponding observed values. This problem has been dramatized by studies of mutations of Val 66 in SNase to ionizable residues and the finding that the corresponding pK_a shifts are much smaller than what are expected from microscopic and from some semimacroscopic estimates. Another related issue is the general process of charge-induced structural changes, which are hard to capture by microscopic simulations. The present work addresses the above challenge by developing an approach that uses the charge of the ionizable residue as a special parameter that drives the partial unfolding process.

The effectiveness of our approach is demonstrated in a study of the pK_a of the Val66Glu mutant of SNase. It is first demonstrated that regular FEP calculations lead to a major overestimation of the observed pK_a and thus basically fail this test case. It is then demonstrated that the overcharging approach leads to a major improvement in the calculated pK_a. The reduced pK_a reflects the protein structural change and water penetration process. Apparently there is a barrier for this partial unfolding process that cannot be surmounted by the regular FEP charging process. Fortunately the overcharging procedure allows us to reach the proper protein configuration where the charged form of the ionizable residue is equilibrated with its surroundings.

It is important to note that the challenge of obtaining correct pK_as in microscopic calculations is, of course, not common to all internal groups. For example, a very recent analysis[26] demonstrated that for Glu35 of lysozyme we can obtain the same results with and without the overcharging approach.

The preliminary results obtained in the present test case indicates that the overcharging approach should have general applications in microscopic studies of electrostatic energies in proteins. This includes studies of pK_as, ion pairs, redox problems, and hydrogen exchange processes.

Acknowledgment. This work was supported by NIH grants 40283 and U19CA105010, and NSF grant MCB-0342276. The computational work was supported by University of Southern California High Performance Computing and Communication Center (HPCC).

References and Notes

(1) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.
(2) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(3) Warshel, A.; Åqvist, J. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 267.
(4) Nakamura, H. *Q. Rev. Biophys.* **1996**, *29*, 1.
(5) Warshel, A.; Papazyan, A. *Curr. Opinion Struct. Biol.* **1998**, *8*, 211.
(6) Simonson, T. *Rep. Prog. Phys.* **2003**, *66*, 737.
(7) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.
(8) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368.

(9) Buono, G. S.; Figueirido, F. E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 85.
(10) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458.
(11) Schutz, C. N.; Warshel, A. *Proteins: Struct., Func., Genet.* **2001**, *44*, 400.
(12) Burykin, A.; Kato, M.; Warshel, A. *Proteins* **2003**, *52*, 412.
(13) Garcia-Moreno, B.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. *Biophys. Chem.* **1997**, *64*, 211.
(14) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; Garcia-Moreno, E. B. *Biophys. J.* **2000**, *79*, 1610.
(15) Fitch, C.; Karp, D.; Gittis, A.; Lattman, E.; Garcia-Moreno, B. *Biophys. J.* **2002**, *82*, 3289.
(16) Damjanovic, A.; Garcia-Moreno, B.; Lattman, E. E.; Garcia, A. E. *Proteins* **2005**, *60*, 433.
(17) Warshel, A. *Biochemistry* **1981**, *20*, 3167.
(18) Valleau, J. P.; Torrie, G. M. *Modern Theoretical Chemistry;* Plenum Press: New York, 1977; Vol. 5.
(19) Kato, M.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 19516.
(20) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.
(21) Chu, Z. T.; Villa, J.; Strajbl, M.; Schutz, C. N.; Shurki, A.; Warshel, A. **2004**.
(22) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.
(23) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640.
(24) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.
(25) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(26) Kato, M.; Pisliakov, A. V.; Warshel, A. *Proteins,* in press.