JCTC Journal of Chemical Theory and Computation

Elucidating the Conformational Dependence of Calculated pK_a Values

Dennis R. Livesay,*[,†,‡] Donald J. Jacobs,[||] Julie Kanjanapangka,[§] Eric Chea,[§] Hector Cortez,[†] Jorge Garcia,[†] Patrick Kidd,[§] Mario Pulido Marquez,[†] Swati Pande,[§] and David Yang[§]

Department of Chemistry, Center for Macromolecular Modeling & Materials Design, and Department of Biological Sciences, California State Polytechnic University, Pomona, California, and Department of Physics and Optical Science, University of North Carolina, Charlotte, North Carolina

Received February 16, 2006

Abstract: The variability within calculated protein residue pK_a values calculated using Poisson−Boltzmann continuum theory with respect to small conformational fluctuations is investigated. As a general rule, sites buried in the protein core have the largest pK_a fluctuations but the least amount of conformational variability; conversely, sites on the protein surface generally have large conformational fluctuations but very small pK_a fluctuations. These results occur because of the heterogeneous or uniform nature of the electrostatic microenvironments at the protein core or surface, respectively. Atypical surface sites with large pK_a fluctuations occur at the interfaces between significant anionic and cationic potentials.

Introduction

Understanding amino acid pK_a fluctuations is key to a deeper understanding of enzyme catalysis.[1] This is especially important considering the dynamic nature of enzyme catalytic site pK_a values. For example, the catalytic Glu169 of the glycolytic enzyme triosephosphate isomerase changes its protonation state four times along its reaction pathway,[2] thus necessitating a dynamic pK_a value. At the beginning of the reaction cycle, Glu169 must be deprotonated (i.e., a low pK_a value) in order for it to act as a general base. Next, the Glu169 pK_a must shift upward such that it can give up the proton to form the enediol intermediate. This protonation/deprotonation cycle is repeated in the second half of the mechanism, finally resulting in the formation of glyceraldehyde-3-phosphate. Previously,[3] we have attributed the first pK_a shift to changes in the local electrostatic environment upon substrate binding. Using Poisson−Boltzmann (PB) continuum theory (described below), the pK_a of Glu169 in the apo structure is calculated to be 0.77, ensuring a deprotonated carboxylate. However, the pK_a is shifted to 8.00 upon substrate binding, making protonation energetically feasible. There are likely two primary factors mediating the remaining three protonation changes. The first, and likely most important, is that changes in the local electrostatics due to the various mechanistic intermediates substantially alter the pK_a of the catalytic site.[4] The second is local conformational changes within the enzyme active site.[5,6] Conformational changes represent a simple way to modulate pK_a values. As a first step toward a computational methodology to probe these complicated acid/base effects, we report the sensitivity of calculated pK_a values to local fluctuations about a native structure.

PB continuum electrostatic theory has become ubiquitous within the computational biology community, see Fogolari et al.[7] for a recent review. One common application of PB theory is in the calculation of pK_a values. There are several similar, yet distinct, PB algorithms for calculating pK_a values using continuum theory, for example, see refs 8−15. However, all are based on the original method of Tanford and Roxby,[16] which assumes that the equilibrium between the acid and base is governed by an *intrinsic* pK_a, where p$K_{a,int}$ is equal to the pK_a if every other titratable site is

* Corresponding author tel.: (909) 869-4409; fax: (909) 869-4434; e-mail: drlivesay@csupomona.edu.

† Department of Chemistry, California State Polytechnic University.

‡ Center for Macromolecular Modeling & Materials Design, California State Polytechnic University.

§ Department of Biological Sciences, California State Polytechnic University.

|| University of North Carolina.

10.1021/ct600066z CCC: $33.50
Published on Web 05/27/2006

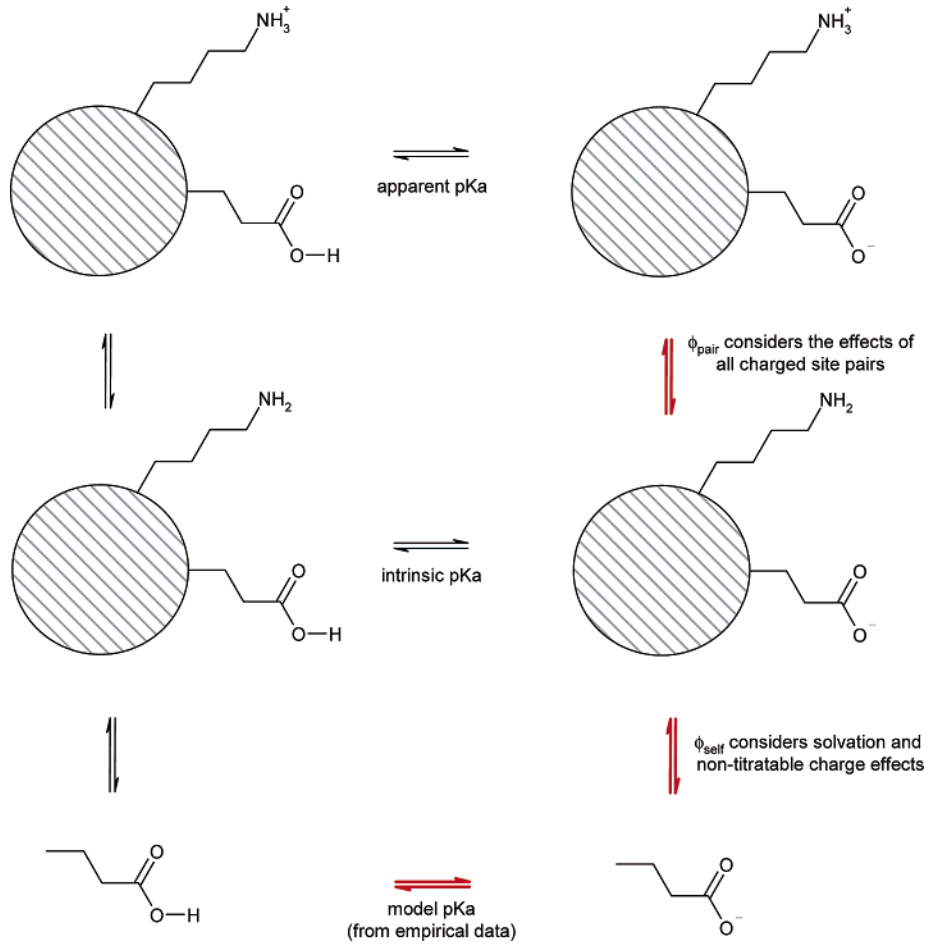

Figure 1. Schematic description of the pK_a calculation algorithm. The method is based on an energy cycle. An *intrinsic* pK_a, which is the hypothetical pK_a for a site if all other titratable sites are neutral, is calculated from the model pK_a by accounting for solvation effects and all nontitratable (partial) charge groups. The apparent pK_a, or real value, is calculated from the intrinsic value by accounting for all charge−charge pair interactions. The *ionic* pK_a is calculated from the model pK_a by ignoring the consequences of Φ_{self}. Despite the schematic shown above, the *apparent* pK_a is actually a mixture of the top two lines because a Boltzmann probability distribution is used to describe the ionization polynomial (see Methods section).

neutral. Common differences within the various pK_a calculation algorithms are related to how flexibility, H-bond networks, and dielectric constants are dealt with.[17] The University of Houston Brownian Dynamics[18] (UHBD) suite of programs calculates the p$K_{a,int}$ from the *model* pK_a, p$K_{a,model}$, which is the experimentally determined aqueous solution pK_a value of the amino acid side chain, by evaluating the effect of nontitratable partial charges and changes in solvation. Computation of the p$K_{a,int}$ requires calculation of the background potential, Φ_{self}, which models the effects of the above considerations. The *apparent* pK_a is calculated from the p$K_{a,int}$ after evaluating the effect of all charge−charge pairs. Each electrostatic potential between two charged sites is calculated by UHBD and is represented as Φ_{pair}. Figure 1 provides a schematic representation of the method; a more detailed description is provided in the Methods section. The approach implemented into UHBD uses a clustering algorithm to reduce the computational expense of evaluating all electrostatic pair potentials in order to compute the actual pK_a.[10,12] The ionization polynomial is exactly solved within a titrating site cluster, whereas a mean-field approximation is used to treat intercluster interactions.

Calculated pK_a values are sensitive to a number of factors, including the chosen interior dielectric constant,[11] H-bond network,[13] and the number of explicit water molecules included.[19] Recently, several reports have focused on understanding the effects of slight conformational changes on calculated pK_a values. For example, a single torsion angle change in hen egg white lysozyme (HEWL) results in large pK_a differences of active site residues.[20] Nielsen and McCammon[17] have investigated the conformational dependence of calculated pK_a values from 41 HEWL X-ray structures, focusing specifically on the ability to correctly identify proton donors and acceptors within two catalytic acids (Glu35 and Asp52). One intriguing conclusion from this work relates to the origins of the conformational dependence of the variability within these two positions. The variability within Glu35 is largely attributed to changes within the set of Φ_{pair}, whereas the variability within Asp52 is caused by changes within both Φ_{self} and Φ_{pair}. Similarly, Kumar and Nussinov have used continuum electrostatics to probe the stability of salt bridges from alternate NMR conformers.[21] Their results indicate that stabilities of salt bridges vary considerably across the conformation ensemble. Moreover, most salt

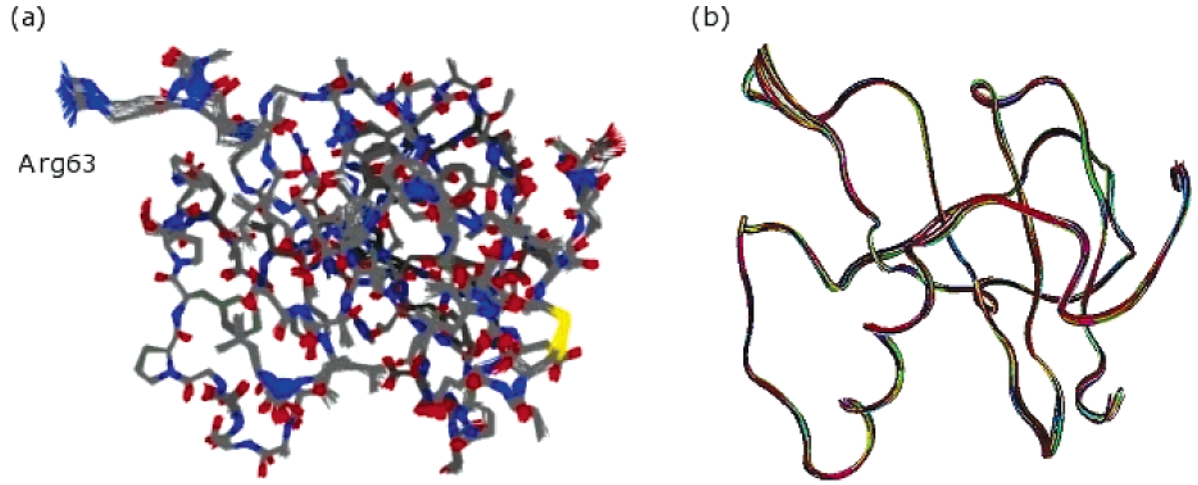

Figure 2. (a) Structural superposition of all RNase Sa conformers investigated. The all-atom RMSD is 0.42 Å. Arg63, which has the greatest structural fluctuations across the ensemble, is highlighted. (b) Backbone superposition, shown in the same orientation as that of part a, of all RNase Sa conformers investigated. The α-carbon RMSD is 0.31 Å.

bridge pairs vary between stabilizing and destabilizing at least once within the population. Changes in salt bridge stability arise because of changes in the location of the charged residues and their orientation (within the salt bridge pair and with respect to other charged sites).

In this report, molecular dynamic (MD) simulations are used to generate an ensemble of protein conformers for three test cases: ribonuclease Sa (RNase Sa) from *Streptomyces aureofaciens*; c-type lysozyme (LYS) from humans, and triosephosphate isomerase (TIM) from *Saccharomyces cerevisiae*. Subsequently, the pK_a values of all titratable sites are calculated and evaluated. Our results indicate that sites buried within the protein core are generally associated with increased pK_a variability. Moreover, we attempt to identify the exact molecular origins of the variability by scrutinizing electrostatic potentials, pK_a values using only Φ_{self} or Φ_{pair}, solvent accessibilities, and titratable site root-mean-square deviations (RMSDs). Finally, it is demonstrated that overall electrostatic free energies, G_{elec}, are generally insensitive to slight conformational changes, especially when compared against the variability within traditional force field potential energy calculations.

Results and Discussion

Variability within pK_a Values. RNase Sa is a small (96 residues) microbial enzyme whose residue pK_a values have been the focus of numerous experimental[22−24] and combined (experimental and theoretical)[25] investigations. Interestingly, the enzyme has 12 acidic residues and only five basic residues. RNase Sa is an ideal starting point for this investigation because of its small size, the fact that several pK_a values of RNase Sa have been solved experimentally, and the fact that a crystal structure is available.[26]

An all-atom structural superposition of the RNase Sa conformers is provided in Figure 2a. The fluctuations are small, as we are purposely investigating small-scale variations. The average pairwise all-atom RMSD is 0.42 Å. Figure 2b provides a backbone superposition of the RNase Sa conformers. The average α-carbon RMSD is 0.31 Å. Significant backbone variability is isolated within the loop region connecting strands β3 and β4. RMSDs describing the structural variability within each titratable residue are provided in Table 1.

Table 1. Rank-Ordered List of All Titratable Averaged pK_a Values, Standard Deviations, Structural Variabilities, and Solvent Accessibilities of RNase Sa[a]

rank order[b]	residue	average pK_a	std. dev.	RMSD[c] (Å)	RSA[d] (Å^2)
1	ASP33	1.47	0.49	0.15	0.5
2	ARG69	15.86	0.40	0.14	2.3
3	ARG65	15.94	0.38	0.14	9.9
4	TYR51	9.20	0.35	0.26	10.4
5	TYR86	8.10	0.32	0.20	11.2
6	TYR80	12.50	0.30	0.22	3.6
7	GLU54	2.49	0.27	0.14	4.2
8	TYR52	10.75	0.21	0.13	1.4
9	TYR55	9.00	0.19	0.28	6.8
10	HIS53	9.48	0.18	0.25	14.0
11	GLU78	4.63	0.18	0.15	6.0
12	ASP79	4.52	0.16	0.14	3.2
13	ASP84	2.84	0.15	0.20	13.7
14	TYR30	7.91	0.14	0.21	19.1
15	TYR81	8.22	0.14	0.20	7.2
16	ARG68	14.22	0.12	0.17	23.1
17	TYR49	6.89	0.12	1.12	43.4
18	GLU14	2.98	0.12	0.13	9.3
19	TRN1	9.49	0.12	0.36	33.0
20	ASP1	2.88	0.10	0.58	33.0
21	TRC96	3.73	0.10	0.27	11.0
22	ASP93	4.09	0.08	0.20	9.9
23	ARG40	12.98	0.07	0.36	47.8
24	GLU74	3.87	0.07	0.19	24.3
25	ASP17	3.98	0.07	0.18	22.1
26	GLU41	3.99	0.06	0.23	30.0
27	HIS85	6.00	0.06	0.37	26.5
28	ASP25	4.48	0.04	0.37	28.4
29	ARG63	12.31	0.03	1.19	58.6
average			0.17	0.29	17.7
std. dev.			0.12	0.26	15.0
correlation[e]				−0.36	−0.63

[a] Average pK_a values and standard deviations are provided for I = 150 mM. Similar deviations are observed at I = 100 and 300 mM. The overall all-atom and α-carbon RMSDs for the structural ensemble are 0.42 and 0.31 Å, respectively. [b] The table is rank-ordered vis-à-vis (largest to smallest) pK_a standard deviation. [c] Titratable atom RMSD. [d] Side-chain solvent accessibility. [e] Linear correlation coefficient between the indicated column and pK_a standard deviation.

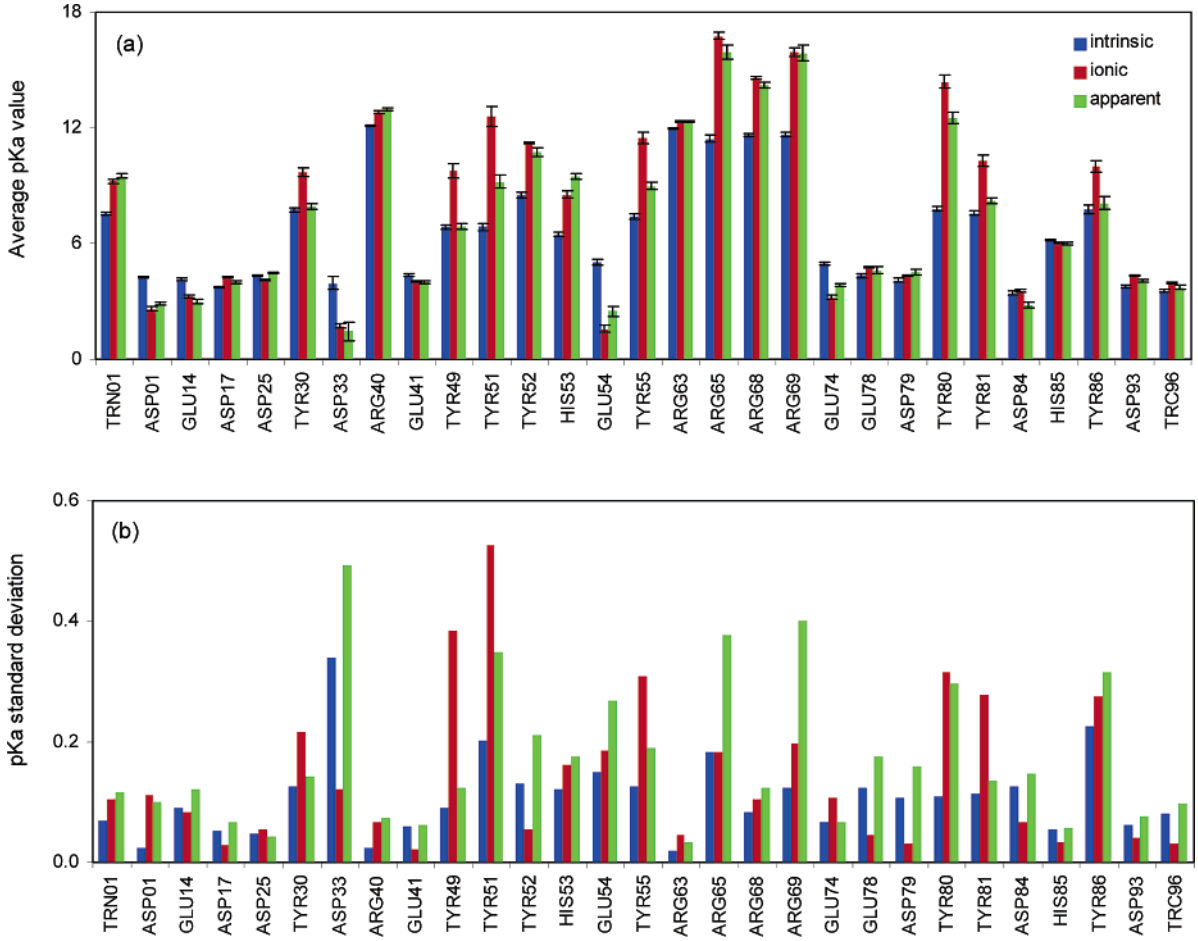

Figure 3. (a) Average intrinsic, ionic, and apparent pK_a values for all sites in RNase Sa ($I = 150$ mM). The intrinsic and ionic pK_a values are calculated by neglecting Φ_{pair} and Φ_{self}, respectively. Error bars represent $\pm$ the standard deviations for each pK_a distribution. Standard deviations are expanded for clarity in part b. Qualitatively similar site-by-site distinctions are observed for $I = 100$ and 300 mM.

Figure 3 provides the average pK_a value and standard deviation for each titratable site. The values provided are for an ionic strength of 150 mM. Values have also been calculated at 100 and 300 mM; however, these results are not shown as their site-to-site distinctions are similar. Table 2 demonstrates that the calculated pK_a values compare favorably to the experimental values.

Table 1, which provides a rank-ordered list of the pK_a standard deviations, shows a wide spectrum of variability within the pK_a fluctuations. Naively, one might expect sites with large conformational fluctuations to also have large pK_a fluctuations. However, this is clearly not the case. In fact, the pK_a fluctuations are slightly anticorrelated with the structural fluctuations, meaning that sites with the smallest conformational fluctuations have the largest pK_a fluctuations. The overall RNase Sa correlation coefficient between the per residue structural variability (calculated as the RMSD for all side-chain target atoms) and the pK_a standard deviation is -0.36 (see Figure 4a). This initially counterintuitive result arises from simple protein structure considerations. Residues on the surface are free to orientate themselves in a variety of ways without drastically affecting their electrostatic surroundings, whereas this is not the case within the crowded protein core. Within the core, slight conformational rearrangements can lead to drastic changes in the electrostatic microenvironments around the buried sites. Figure 4b plots side-chain atomic solvent accessibility against the pK_a variability. A similar negative correlation between the solvent accessibility and pK_a variability is observed in the LYS and TIM examples. Curiously, no significant correlation is observed between the pK_a fluctuations and fluctuations within the overall potential energy values (calculated using the CHARMM[27] force field). The range of pK_a/potential energy correlations for the 29 different RNase Sa titratable sites is $\{-0.22; 0.25\}$. Moreover, the correlation between the force field potential energy and the electrostatic free energy (G_{elec}) at pH 7.0 is also insignificant ($R = -0.20$). This result is discussed in more detail below.

The above points are exemplified by a single RNase Sa arginine pair. For example, the pK_a variability within Arg65 is quite high (standard deviation = 0.38), whereas the variability within Arg63 is the smallest (standard deviation = 0.03). Figure 5a compares the structural superposition of each residue's conformers. However, the structural variability within the guanidinium group of Arg65 (CZ RMSD = 0.14 Å) is much smaller than that of Arg63 (CZ RMSD = 1.18 Å). Arg65 is buried (side chain ASA = 9.92 Å^2) within the core, which significantly reduces its conformational freedom. Nevertheless, because of the heterogeneous nature of the electrostatic environment within the core, the slight confor-

Table 2. Comparison of Calculated and Experimental pK_a Values for RNase Sa[a]

	experiment $I = 100$ mM	calculated $I = 100$ mM	calculated $I = 150$ mM	calculated $I = 300$ mM
TRN1	9.14	9.83	9.49	9.33
ASP1	3.44	3.17	2.88	2.93
GLU14	5.02	3.94	2.98	3.06
ASP17	3.72	4.44	3.98	3.99
ASP25	4.87	4.83	4.48	4.43
TYR30	11.3	11.85	7.91	7.92
ASP33	2.39	2.47	1.47	1.60
GLU41	4.14	4.40	3.99	4.03
TYR49	10.6	10.58	6.89	6.88
HIS53	8.27	10.18	9.48	9.21
GLU54	3.42	5.14	2.49	2.61
GLU74	3.47	4.61	3.87	3.88
GLU78	3.13	7.65	4.63	4.60
ASP79	7.37	5.62	4.52	4.47
ASP84	3.01	3.49	2.84	2.94
HIS85	6.35	6.83	6.00	5.98
ASP93	3.09	4.50	4.09	4.05
TRC96	2.42	2.97	3.73	3.72

[a] Calculated pK_a values are the average of all conformers using an interior (protein) dielectric of 20, and an exterior (solvent) dielectric of 80, at three different ionic strengths. Experimental values are taken from Laurents et al.[25] The correlation coefficient computed from the ensemble-averaged pK_a ($R = 0.90$) is similar to, albeit slightly less ($R = 0.93$) than, the correlation coefficient of the theoretical results reported in Laurents et al. Correlation coefficients are only computed for the 100 mM results, which is the same as the experimental conditions.

mational changes in Arg65 can have pronounced effects on its pK_a value. Conversely, Arg63 is completely solvent-exposed (side chain ASA = 58.57 Å^2) and is, thus, able to explore a much larger conformational space. Because the electrostatic environment on the RNase Sa surface is more uniform, at least compared to the myriad electrostatic microenvironments within the protein core, the large conformational changes with Arg63 have little effect on its calculated pK_a value.

Arg65 is part of an electrostatic network that includes Asp33, Glu54, and Arg69 (Figure 5b). All three also have large fluctuations within their pK_a value distributions. In fact, the variability within Asp33 is the largest for RNase Sa. As one site is perturbed, there is a local change in the electrostatic microenvironment that affects all four pK_a values simultaneously. Comparable sensitivities are observed in other buried charged clusters. The variabilities within the LYS and TIM pK_a values are similar to the RNase Sa results. Moreover, the inverse correlation between ASA and RMSD is also qualitatively similar (see Tables 3 and 4). However, some interesting deviations to the overall trends, which are also discussed in the next section, do occur in TIM.

pK_a Variability within Solvent-Exposed Sites. It is demonstrated above that the extent of pK_a variability can generally be ascribed to solvent accessibility and structural variability, which are, of course, related. Like Arg63 of

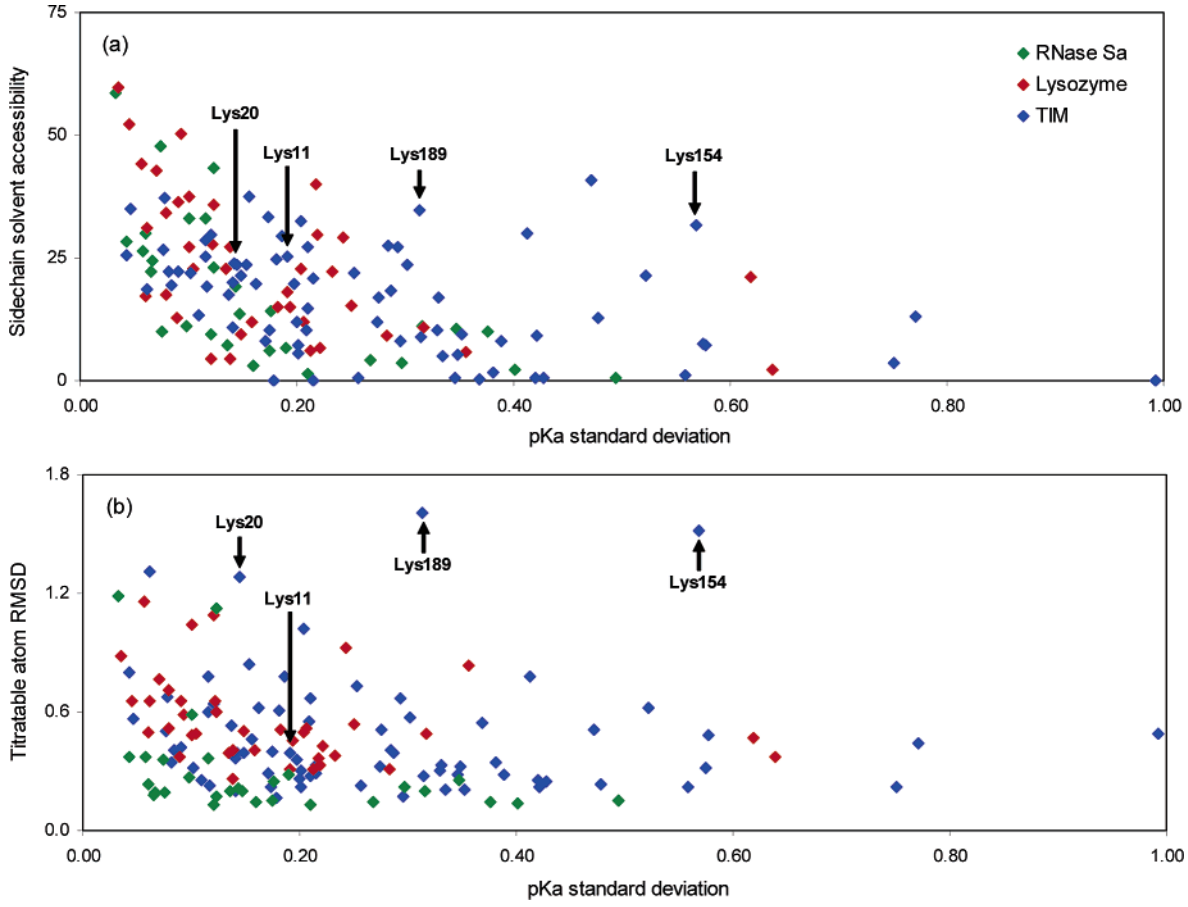

Figure 4. (a) Side-chain solvent accessibility vs the standard deviation for each pK_a distribution for the three investigated proteins. Note the nonlinear dependence of the solvent accessibility effects. (b) Structural RMSD for each titratable target atom vs the standard deviation for each pK_a distribution. A similar nonlinear correlation is observed. In both figures, the four TIM lysines discussed in the text are highlighted.

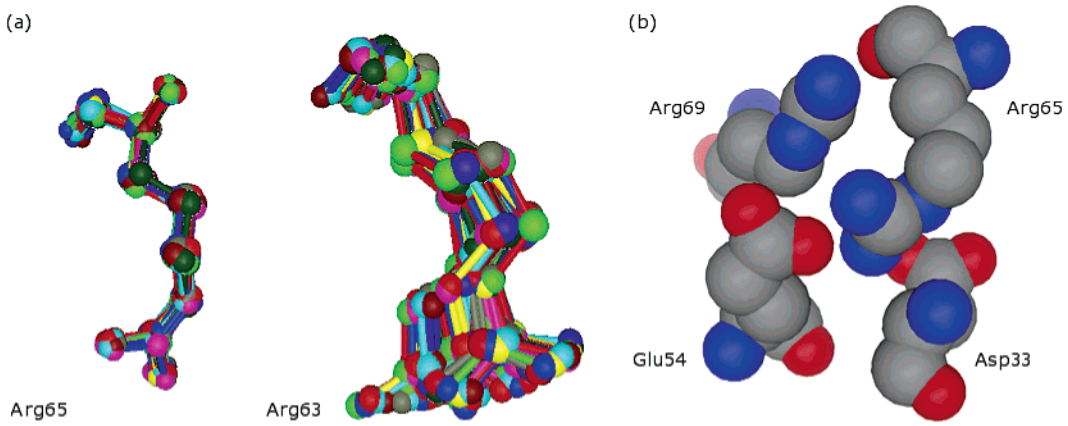

Figure 5. (a) Structural superposition of RNase Sa's Arg65 and Arg63. The structural variability within the solvent-exposed Arg63 is the largest of all RNase Sa residues. However, because the electrostatic microenvironment of solvent-exposed sites is generally uniform, its pK_a variability is quite small. On the other hand, the structural variability within the buried Arg65 is small, yet its pK_a variability is quite large—the third largest of all RNase Sa residues. (b) The buried Arg65 is sandwiched between Asp33, Glu54, and Arg69. As a consequence, small conformational fluctuations can significantly affect the electrostatic microenvironment of these residues. In fact, residues Asp33, Arg69, and Arg65 have the three largest pK_a value standard deviations (see Table 1). Glu54, ranked seventh, also displays significant pK_a variability.

RNase Sa, Lys20 of TIM typifies the normal situation of a solvent-accessible site (see Figure 6a). Despite the general consistency of these trends, there are some notable exceptions. It can be seen in Figure 4 that plotting RMSD or ASA versus pK_a variability loosely approximates a quadratic curve in all three examples, meaning sites with large pK_a fluctuations are almost always buried (or conformationally constrained). The most egregious exceptions to this trend occur in TIM. For example, Lys154 and Lys189, which are both solvent-accessible, have atypically large pK_a fluctuations. The basis of the observed pK_a variability is revealed in Figure 6. The electrostatic potential maps in Figure 6 clearly indicate that the positions of Lys154 and Lys189 are located at interfaces between significant anionic and cationic regions. As the position of the titratable atom fluctuates, its electrostatic microenvironment also changes. Consequently, a large range of pK_a values is observed for these sites. Normally, the distinctions in the electrostatic microenvironments on the protein surface are not very significant; however, in these two cases, they clearly are. The pK_a ranges for Lys154 and Lys189 are {10.68; 12.45} and {10.74; 12.29}, respectively.

Across all three protein examples, solvent-exposed sites with constrained conformational variability can occur. In all cases analyzed, the dynamics within these sites are constrained by some sort of noncovalent interaction. For example, Lys11, which has a solvent accessibility quantitatively similar to that of Lys20, is highly constrained because of an extended electrostatic network within the active-site region.[3] Lys11 is most strongly interacting with Glu96 via a strong ionic bond. This salt bridge constricts the dynamics of Lys11. Lys11 is also electrostatically interacting with the catalytic Glu164, His94, Tyr100, and Cys125 (Figure 6d). The pK_a variability within Lys11 is ranked in the middle third of all TIM sites. The bulk of the sites with pK_a fluctuations of similar scale are inaccessible to solvents.

Elucidating the Origins of pK_a Variability within the Core. As discussed above (see Figure 1), the standard procedure of computing pK_a values uses a two-step process. The first step calculates the *intrinsic* pK_a from the *model* pK_a by accounting (via Φ_{self}) for the solvation and background charge changes that occur when going from a fully solvated side chain to the hypothetical neutral protein environment. The *apparent* pK_a, which is the final calculated value, is calculated from the intrinsic pK_a by accounting (via Φ_{pair}) for a more realistic charged-protein environment. In this step, the electrostatic potential between all titratable charge pairs is evaluated. In RNase Sa, the largest pK_a variability occurs in Asp33. Figure 7 plots the difference between all electrostatic potentials between the two RNase Sa conformers with the most extreme Asp33 pK_a values. Curiously, there is significant and consistent variability within Φ_{self} (on diagonal), whereas the variability within Φ_{pair} is more intermittent. Moreover, the difference within the self-potential of Asp33 (1.3 kcal/mol/*e*) dwarfs all other differences (the second largest difference is 0.3 kcal/mol/*e*).

To better understand the effects of potential variability on calculated pK_a values, we compute pK_a values using only one of the two steps from the normal procedure. Values calculated using only Φ_{self} are deemed *intrinsic* pK_a's, whereas values calculated using only Φ_{pair} are called *ionic* pK_a's. These values are also presented in Figure 3a alongside the *apparent* pK_a values. The difference between the three different "pK_a values" is small for most solvent-exposed sites (e.g., Arg40, Arg63, and His85). However, large differences are common within buried sites. There is a slight negative correlation ($R = -0.34$) between site accessibility and the ΔpK_a (defined as $|pK_{aintrinsic} - pK_{aionic}|$). The ionic p$K_a$ values are generally closer to the apparent values than the intrinsic values, which highlights the increased importance of the various formal charges within the protein. This result is especially true for sites that are largely inaccessible to solvents.

In all but three sites, the intrinsic pK_a is calculated to be less than the ionic pK_a. Two of the exceptions correspond to Asp33 and Glu54, both of which are discussed above. Figure 3b expands the standard deviations observed within

Table 3. Rank-Ordered List of All Titratable Averaged pK_a Values, Standard Deviations, Structural Variabilities, and Solvent Accessibilities of LYS[a]

rank order[b]	residue	average pK_a	std. dev.	RMSD[c] (Å)	RSA[d] (Å^2)
1	ASP67	3.43	0.64	0.37	2.3
2	LYS69	15.16	0.62	0.47	21.1
3	TYR54	8.65	0.36	0.83	5.9
4	ARG62	15.01	0.32	0.49	10.9
5	TYR38	7.78	0.28	0.31	9.1
6	GLU7	2.53	0.25	0.53	15.3
7	TYR45	7.51	0.24	0.92	29.0
8	LYS13	12.60	0.23	0.38	22.2
9	TYR124	7.39	0.22	0.43	6.7
10	ARG101	13.34	0.22	0.33	29.7
11	ARG10	13.18	0.22	0.37	39.9
12	ASP53	3.64	0.21	0.31	6.1
13	ASP91	3.67	0.21	0.52	12.1
14	LYS1	12.88	0.20	0.50	22.7
15	ASP49	2.29	0.19	0.46	15.0
16	ARG98	14.07	0.19	0.31	18.0
17	TYR20	6.43	0.18	0.51	14.9
18	LYS97	11.38	0.16	0.41	12.0
19	ASP18	2.38	0.15	0.51	9.4
20	GLU35	5.17	0.14	0.26	4.4
21	ARG21	13.06	0.14	0.41	27.1
22	TRN1	8.28	0.13	0.39	22.7
23	TYR63	7.14	0.12	0.60	35.8
24	ARG5	13.73	0.12	0.65	27.7
25	TRC130	2.43	0.12	1.09	4.4
26	ASP87	2.65	0.10	0.49	22.8
27	ARG119	13.24	0.10	0.48	27.3
28	HIS78	6.09	0.10	1.04	37.6
29	ARG122	13.49	0.09	0.59	50.4
30	ARG115	12.68	0.09	0.65	36.3
31	ASP102	2.35	0.09	0.37	12.9
32	LYS33	11.49	0.08	0.71	17.4
33	ARG113	13.11	0.08	0.51	34.2
34	ARG107	12.04	0.07	0.77	42.9
35	GLU4	3.93	0.06	0.66	31.1
36	ASP120	2.81	0.06	0.50	17.1
37	ARG50	12.71	0.06	1.16	44.2
38	ARG41	12.79	0.05	0.65	52.3
39	ARG14	12.38	0.03	0.88	59.8
average			0.18	0.56	23.4
std. dev.			0.13	0.22	14.6
correlation[e]				−0.30	−0.48

[a] Average pK_a values and standard deviations are provided for I = 150 mM. Similar deviations are observed at I = 300 mM. The overall all-atom and α-carbon RMSDs for the structural ensemble are 0.73 and 0.58 Å, respectively. [b] The table is rank-ordered vis-à-vis (largest to smallest) pK_a standard deviation. [c] Titratable atom RMSD. [d] Side-chain solvent accessibility. [e] Linear correlation coefficient between the indicated column and pK_a standard deviation.

all calculated pK_a values in order to facilitate comparisons. As suggested by Figure 7, Asp33 is unique because of its large variability within its intrinsic pK_a distribution. This result indicates that Asp33 is very sensitive to local fluctuations within the background electrostatics. To explore this result more closely, correlations between the intrinsic pK_a values for all titratable site pairs are computed (data not shown). The site most strongly correlated with Asp33 is Arg65, which suggests that the relative location of these two sites has a pronounced effect on the background electrostatics. This initially counterintuitive result (one might expect variability between two interacting *charged* residues to affect the *ionic* pK_a more than the intrinsic pK_a) is explained by the fact that the carboxylate of Asp33 is doubly hydrogen-bonded to the nontitrating NE and NH1 atoms of Arg65. As a consequence, a slight structural rearrangement between the two significantly affects the local background electrostatics (as exemplified in Figure 7). Changes in the protonation or deprotonation state of Asp33 or Arg65 has no effect on the presence of the two hydrogen bonds, which explains the reduced ionic pK_a correlation for this pair. Because of their structural proximity, significant fluctuations are observed within $\Phi_{Asp33-Arg65}$ and $\Phi_{Asp33-Arg69}$. However, large fluctuations are not observed in any other Asp33 site pairs, which keeps its p$K_{a,ionic}$ from varying significantly. While the correlation between the contacting Asp33−Arg65 pair is the strongest observed for Asp33, it should be noted that several structurally remote sites are also strongly correlated with it. The origin of these correlations is unclear. Future work will attempt to identify their origin.

Sites with the largest variability within their ionic pK_a values are generally tyrosines. As can be seen in Figure 3b, Tyr51 is identified as the RNase Sa site with the most significant ionic pK_a variability. Figure 7a reveals that significant changes within the pairwise potentials occur within $\Phi_{Tyr51-Glu74}$, $\Phi_{Tyr51-Glu78}$, and $\Phi_{Tyr51-Tyr80}$, which result in the ionic pK_a fluctuations. Figure 7b demonstrates that these four sites constitute a second electrostatic tetrad (distinct from the Asp33−Glu54−Arg65−Arg69 tetrad discussed above). In this charge cluster, slight conformational changes significantly affect ionic pK_a values. Nevertheless, a significant fraction of the apparent pK_a variability within Glu74 and Glu78 is also attributed to fluctuations within the intrinsic pK_a (Figure 3b). This point illustrates one of the main results of this investigation, that being apparent pK_a fluctuations within the protein core can arise from changes in both the background and pairwise electrostatic interactions. Apparent pK_a fluctuations that arise from convolutions of Φ_{self} and Φ_{pair} are also frequently observed in LYS and TIM. As mentioned previously, Nielsen and McCammon[17] report an identical conclusion regarding the origins of the variability within Asp52 from their comparison of 41 HEWL X-ray structures.

Variability within Overall G_{elec} Values. G_{elec}, which is also calculated by UHBD as part of the pK_a calculation, is the purely electrostatic portion of the overall protein free energy. (A brief description of how G_{elec} is determined is provided in Livesay et al.[28]) Because G_{elec} is frequently used to assess the electrostatic portions of molecular recognition events[28−31] and overall protein stability,[32,33] understanding the conformational sensitivity of this quantity is also paramount. For all three proteins, it is found that the average snapshot-to-snapshot $\Delta G_{elec} \approx 0$, meaning that the stabilizing and destabilizing changes tend to cancel each other out. Table 5 lists the average snapshot-to-snapshot absolute value of ΔG_{elec} for the three protein examples; the standard deviation of $|\Delta G_{elec}|$ and its overall range is also provided. As with

Table 4. Rank-Ordered List of All Titratable Averaged pK_a Values, Standard Deviations, Structural Variabilities, and Solvent Accessibilities of TIM[a]

rank order[b]	residue	average pK_a	std. dev.	RMSD[c] (Å)	RSA[d] (Å^2)	rank order[b]	residue	average pK_a	std. dev.	RMSD[c] (Å)	RSA[d] (Å^2)
1	TYR207	16.77	0.99	0.49	0.0	38	LYS113	12.07	0.21	0.67	27.3
2	TYR48	11.17	0.77	0.44	13.1	39	ASP80	2.53	0.21	0.55	10.2
3	GLU103	1.63	0.75	0.22	3.7	40	LYS134	11.57	0.20	1.02	32.5
4	TYR45	10.77	0.58	0.49	7.2	41	ARG204	15.88	0.20	0.31	7.3
5	TYR66	11.20	0.58	0.32	7.5	42	ASP179	2.42	0.20	0.22	5.7
6	LYS154	11.70	0.57	1.52	31.7	43	GLU132	2.79	0.20	0.26	11.9
7	TYR163	19.49	0.56	0.22	1.1	44	GLU202	2.70	0.20	0.36	19.7
8	ASP197	2.36	0.52	0.62	21.3	45	LYS11	12.42	0.19	0.39	25.2
9	GLU76	1.86	0.48	0.24	12.9	46	GLU131	2.80	0.19	0.78	29.5
10	TYR100	10.09	0.47	0.51	40.9	47	LYS220	11.84	0.18	0.60	24.8
11	LYS111	14.94	0.43	0.25	0.5	48	GLU128	5.73	0.18	0.17	0.1
12	ARG97	14.67	0.42	0.22	9.3	49	GLU238	2.65	0.18	0.40	10.3
13	ASP226	−0.41	0.42	0.25	0.6	50	ARG144	14.03	0.17	0.22	33.4
14	LYS16	11.05	0.41	0.78	29.9	51	GLU96	1.32	0.17	0.29	8.0
15	ASP110	2.83	0.39	0.29	8.0	52	LYS194	11.14	0.16	0.62	19.8
16	HIS94	6.15	0.38	0.35	1.7	53	LYS198	11.96	0.16	0.46	37.6
17	CYS40	11.08	0.37	0.55	0.2	54	LYS68	10.53	0.15	0.84	23.5
18	ARG188	17.67	0.35	0.21	9.4	55	ASP221	2.05	0.15	0.40	21.5
19	ASP224	2.39	0.35	0.32	5.3	56	LYS20	11.49	0.14	1.28	23.6
20	CYS125	15.06	0.35	0.28	0.5	57	GLU21	3.16	0.14	0.39	23.8
21	ASP105	0.42	0.33	0.21	5.0	58	GLU143	4.57	0.14	0.20	10.8
22	ASP140	2.57	0.33	0.33	17.1	59	LYS55	11.71	0.14	0.36	20.1
23	ARG2	14.46	0.33	0.30	10.2	60	LYS222	11.84	0.14	0.53	17.5
24	GLU36	3.26	0.32	0.28	8.8	61	LYS88	11.21	0.12	0.64	29.8
25	LYS189	11.14	0.31	1.61	34.7	62	ASP104	3.09	0.12	0.23	19.2
26	ASP47	3.11	0.30	0.57	23.5	63	LYS137	12.01	0.12	0.60	25.2
27	ARG98	17.62	0.30	0.18	8.0	64	LYS236	11.50	0.12	0.78	28.7
28	LYS54	12.00	0.29	0.67	27.1	65	ARG25	13.96	0.11	0.26	13.4
29	LYS133	12.90	0.29	0.39	18.4	66	ASP241	3.86	0.10	0.32	22.1
30	LYS106	13.37	0.28	0.41	27.6	67	ASP182	3.94	0.09	0.42	22.2
31	GLU24	3.18	0.28	0.51	16.8	68	TRC247	3.48	0.08	0.40	19.5
32	GLU152	2.32	0.27	0.32	11.8	69	ASP155	2.89	0.08	0.35	22.3
33	GLU164	−0.41	0.26	0.23	0.7	70	HIS102	6.78	0.08	0.67	37.1
34	LYS83	12.07	0.25	0.73	22.0	71	GLU151	4.28	0.08	0.50	26.7
35	HIS184	7.31	0.21	0.29	0.0	72	TRN1	7.79	0.06	1.31	18.7
36	GLU178	2.79	0.21	0.33	20.7	73	GLU33	4.24	0.05	0.56	34.9
37	ARG246	14.63	0.21	0.28	14.6	74	ASP84	4.14	0.04	0.80	25.4
average			0.27	0.48	17.0						
std. dev.			0.18	0.30	10.9						
correlation[5]				−0.09	−0.40						

[a] Average pK_a values and standard deviations are provided for $I = 150$ mM. Similar deviations are observed at $I = 300$ mM. The overall all-atom and α-carbon RMSDs for the structural ensemble are 0.59 and 0.46 Å, respectively. [b] The table is rank-ordered vis-à-vis (largest to smallest) pK_a standard deviation. [c] Titratable atom RMSD. [d] Side-chain solvent accessibility. [e] Linear correlation coefficient (for all 74 titrable sites) between the indicated column and pK_a standard deviation.

the pK_a variations, the G_{elec} fluctuations within RNase Sa are the smallest ($\langle|\Delta G_{elec}|\rangle = 0.35$ kcal/mol) of the three examples investigated. Somewhat surprisingly, the variation with the UHBD G_{elec} values is uncorrelated with the CHARMM potential energy values (see Figure 8). The lack of correlation arises from the reduced variability within the G_{elec} values. For example, the standard deviation within the UHBD G_{elec} values is 10% of the average value, whereas it is 74% of the average CHARMM potential energy value. From this result, it can be inferred that G_{elec} is fairly insensitive to slight conformational changes, especially when compared to traditional force field methods. Moreover, the observed robustness within G_{elec} strengthens the conclusions of studies that use G_{elec} to probe single protein conformations, such as those referenced above.

Conclusions

Our ultimate goal is to develop a robust computational framework to understand pK_a changes along a reaction coordinate. In this report, we use MD simulations to generate a conformational ensemble within three protein examples (RNase Sa, TIM, and LYS) to determine the conformational sensitivity of calculated pK_a values. The conformational variability is explicitly designed to be small in order to focus this investigation on the effects of *slight conformational changes*. These results provide a baseline of pK_a fluctuations

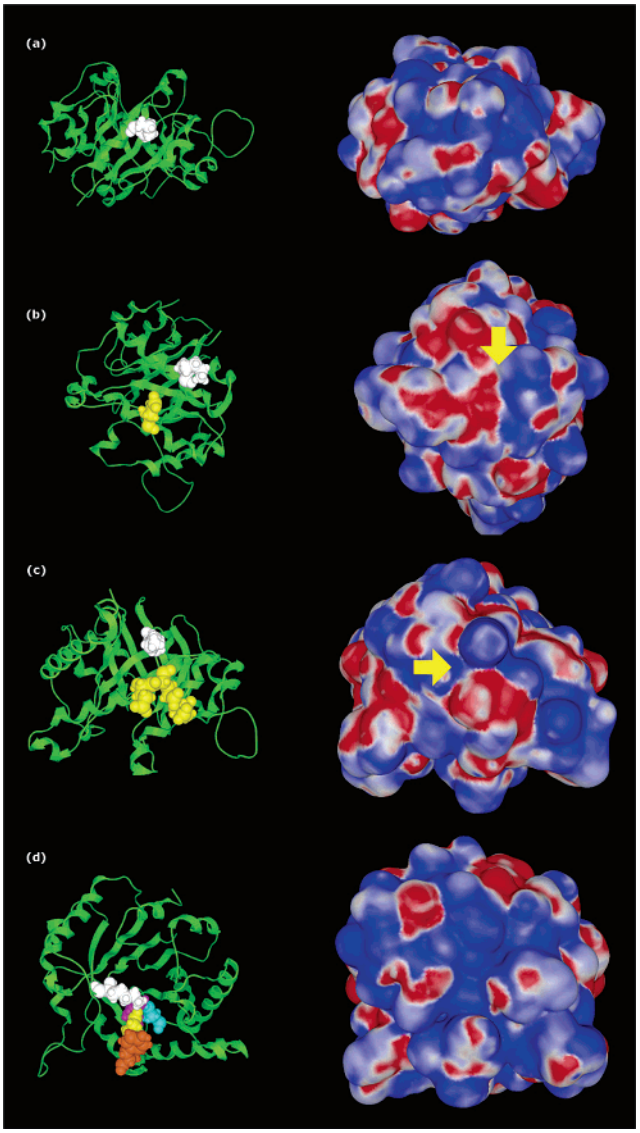

Figure 6. Comparison of the electrostatic environment around four TIM lysine residues. Structural cartoons, oriented the same as the electrostatic potential maps, are provided to facilitate comparisons. The target lysine residues are colored white, and are (from top to bottom) Lys20, Lys189, Lys154, and Lys11. (a) Lys20 typifies the normal case where solvent-exposed residues exhibit large structural fluctuations, yet their pK_a variability is small. This result occurs because the electrostatic microenvironment of the surface is more uniform than in the core; the electrostatic environment surrounding Lys20 is predominantly cationic, with a few small anionic regions. The solvent-exposed (b) Lys189 and (c) Lys154 are atypical because these sites have significant pK_a variability (see Figure 4). This result occurs because there are stark anionic/cationic electrostatic potential regions near these sites. The culprit anionic/cationic potential interfaces are highlighted by the yellow arrow. Acidic residues that predominantly define the anionic regions near the two lysine residues are colored yellow. (d) The solvent-exposed Lys11 is also atypical because its structural variability is significantly constrained. The constrained structural variability within Lys11 is due to a strong salt bridge between it and Glu96 (colored yellow). Also displayed are Glu164 (cyan), His94 (violet), and Tyr100 (orange), which make up an extended electrostatic network at the active site of the enzyme.

that can be used in subsequent investigations. Future work will attempt to incorporate protein flexibility and changes in electrostatic environment due to substrates and reaction intermediates, similar to our previous work,[3] to better model these effects.

Our results indicate that sites buried in the protein core are very sensitive to slight structural fluctuations, whereas sites on the surface are generally robust. A few exceptions to the latter trend are observed in TIM, which can be explained by their proximity to drastic changes in the anionic/cationic character of the electrostatic potential surfaces. In summary, the results presented herein (for both buried and exposed sites) highlight the structural sensitivity of calculated pK_a values within heterogeneous electrostatic environments. Heterogeneity within the local electrostatic environment is usually associated with the crowded protein core; however, as demonstrated by TIM, it can also be significant on the protein surface. Finally, overall G_{elec} values are generally robust to slight conformational changes. This final result is especially apparent when compared against the increased variability within traditional force field techniques.

Separating the apparent pK_a calculation into its intrinsic pK_a and ionic pK_a constituent parts indicates that the observed pK_a variability arises from effects associated with both nontitratable and titratable charges. For example, in the case of RNase Sa, Asp33 is hydrogen-bonded to the nontitrating NE and NH1 atoms of Arg65. As a consequence, much of the variability within Asp33, which has the large apparent pK_a variability of all RNase Sa sites, is due to nontitratable (background) charges. Conversely, slight conformational fluctuations have a more significant effect on the pairwise electrostatic potentials of Tyr51 than its self-potential. Similar results are observed in LYS and TIM.

Methods

Calculation of pK_a Values. Titratable residue pK_a values are calculated using the UHBD suite of programs.[34] All calculations employ the same approach that we have reported previously.[3,28,31,32,35] In the approach, Φ_{self} is used to calculate the *intrinsic* pK_a from the model values. When calculating Φ_{pair}, all background charges are set to zero, because they are already included in the *intrinsic* pK_a. With the Φ_{self} and Φ_{pair} potentials in hand, the pK_a is determined after considering all possible ionization states, meaning that, despite the schematic shown in Figure 1, the *apparent* pK_a is actually a mixture of the top two lines. For example, Figure 1 encapsulates four different ionization states: Lys^{+1}/Glu^{1-}, Lys^{1+}/Glu^{0}, Lys^{0}/Glu^{1-}, and Lys^{0}/Glu^{0}. A Boltzmann probability distribution is used to describe each possible ionization state. Over a series of pH values, the fractional charge of each site is calculated as the sum of the probabilities when ionized. From the Henderson−Hasselbalch equation, the pK_a is simply defined as the pH at which the fractional charge is ±0.5, for acids and bases, respectively. Because sites can be either neutral or ionized, it also follows that both sides of the horizontal equilibria are evaluated when the pK_a values are determined. For large numbers of titratable sites, the computational cost of considering all 2^N possible ionization states is prohibitive. To make the problem computationally

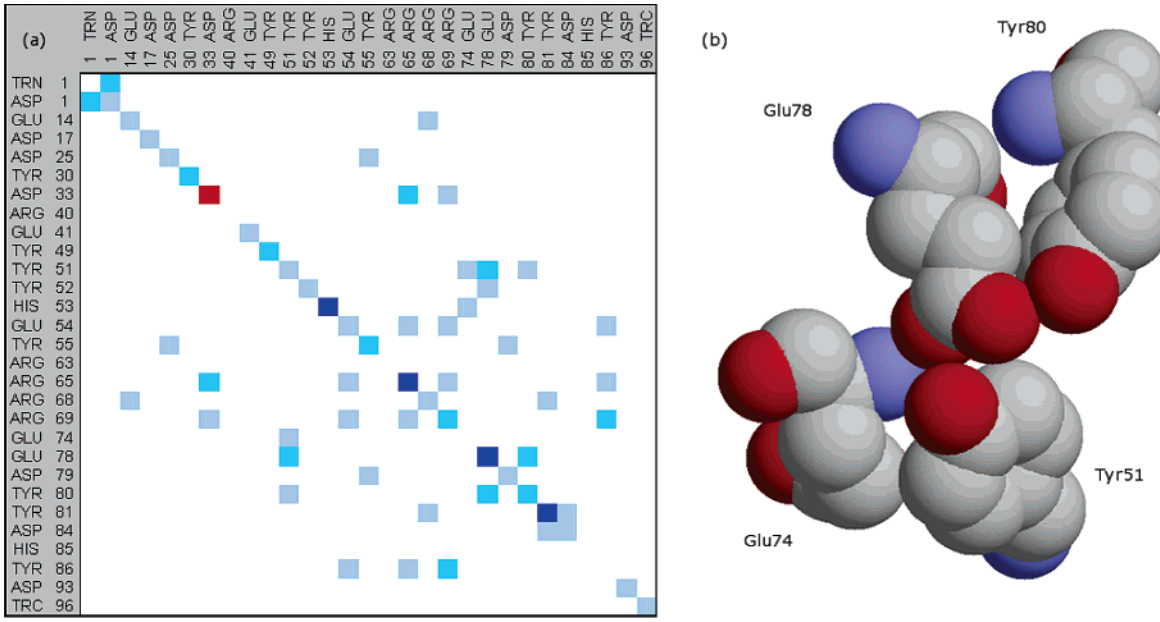

Figure 7. (a) Differences in electrostatic potentials between the RNase Sa conformer pair with the most extreme Asp33 pK_a values. Off-diagonal values correspond to Φ_{pair}, whereas on-diagonal values correspond to Φ_{self}. The three shades of blue (light to dark) correspond to differences of 0.1 kcal/mol/*e*, 0.2 kcal/mol/*e*, and 0.3 kcal/mol/*e*; red corresponds to a difference of 1.3 kcal/mol/*e*. (b) Tyr51 has the most significant ionic pK_a variability (see Figure 3b). For this site, the variability arises from changes within $\Phi_{Tyr51-Glu74}$, $\Phi_{Tyr51-Glu78}$, and $\Phi_{Tyr51-Tyr80}$, which constitute a tight, solvent-exposed cluster of four titratable sites.

Table 5. $|\Delta G_{elec}|$ Variability Statistics[a]

protein	$\langle\|\Delta G_{elec}\|\rangle$ (kcal/mol)	std. dev. (kcal/mol)	minimum (kcal/mol)	maximum (kcal/mol)
RNase Sa	0.35	0.25	0.00	1.59
LYS	0.90	0.67	0.01	2.56
TIM	1.08	0.92	0.06	4.72

[a] Statistics describing the distribution of contiguous snapshot-to-snapshot $|\Delta G_{elec}|$ values.

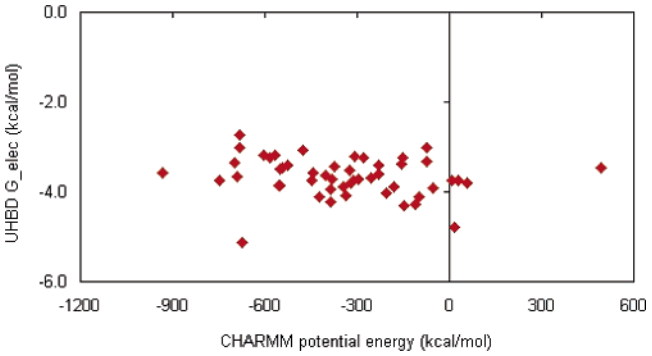

Figure 8. No obvious correlation exists between the UHBD-calculated G_{elec} values of the RNase Sa conformers and the corresponding potential energies computed from the CHARMM force field ($R = -0.20$). This result arises from the general lack of conformational sensitivity within the G_{elec} values. As a consequence, G_{elec} is determined to be rather insensitive to small structural fluctuations.

tractable, UHBD calculates pK_a values using the cluster method described in Gilson[12] and Antosiewicz et al.[10] The *ionic* pK_a is simply calculated from the model value by setting all Φ_{self} values to zero.

The linear Poisson−Boltzmann equation (LPBE) is solved using the Choleski preconditioned conjugate gradient method. The LPBE is used, versus the computationally more expensive nonlinear Poisson−Boltzmann equation (NLPBE), because of the large number of electrostatic potential calculations required to calculate all pK_a values within a given protein structure. The protein is centered on a 65 × 65 × 65 grid with each grid unit equaling 1.5 Å. Focusing is used around each titrating site with the grid spacing becoming 1.2, 0.75, and 0.25 Å. In all calculations, a solvent dielectric constant of 80 and a protein dielectric constant of 20 are used. Protein partial charges are taken from the CHARMM parameter set[27] and radii from the Optimized Potentials for Liquid Systems.[36] The ionic strength varies between 100 and 300 mM, and the temperature is 298 K. Intrinsic and ionic pK_a values are calculated using the standard procedure, but without including the background and charge pair effects, respectively.

Electrostatic Potential Maps. Electrostatic potential maps are calculated using the NLPBE solver within the Molecular Operating Environment (MOE) software package. The proteins are centered on a 33 × 33 × 33 grid. A solvent dielectric constant of 80 is used, with a protein dielectric constant of 4, which are standard values in electrostatic potential map calculations.[37] Electrostatic potential maps calculated using an interior dielectric constant of 20 are qualitatively similar (results not shown). Protein partial charges are taken from the CHARMM parameter set.[27] The temperature is 298 K; the ionic strength is 150 mM, and the protein concentration is 0.001 M. Electrostatic potentials are rendered in blue and red onto the protein solvent-accessible surface at ±1.0 kcal/mol/*e*, respectively. Electrostatic potential maps are provided for only one exemplar conformer; however, all qualitative conclusions based on that exemplar are robust to structural variations.

Protein Structures and Molecular Dynamics. Protein structures are modified versions of the coordinates retrieved

from the Protein Databank (PDB). The continuum electrostatics method implemented in the UHBD suite of programs requires explicit polar hydrogen atoms, which are added using the MOE software package. Proteins (and PDB identification codes) for the protein structures used are RNase Sa from *S. aureofaciens* (1RGG),[26] triosephosphate isomerase from *S. cerevisiae* (7TIM),[38] and human c-type lysozyme (1JSF).[39] Canonical ensemble (fixed *NVT*) in vacuo molecular dynamics simulations, as implemented in MOE, are used to generate the ensemble of conformers. In each example, the time scale of the simulations is 1 ns, and the time step is 0.001 ps. A steepest-descent minimization (till convergence) and an equilibration phase (1000 iterations) precede the sampling phase of the simulation. In the cases of RNase Sa, conformers are sampled uniformly over a narrow 100 ps range (sampled every 1.5 ps) to specifically focus on slight fluctuations. The structural variability within the nonsampled phase of the simulation suggests that its pK_a variability should be consistent with the sampled conformations. In the case of TIM and LYS, conformers are uniformly sampled every 40 and 50 ps, respectively, over the entire MD simulation. Despite the reduced sampling time of the RNase Sa simulation, Figure 4a clearly indicates that the scale of the RNase Sa structural variations is in line with the other two examples. It is obvious that this in vacuo simulation protocol is unacceptable to determine realistic aqueous-phase dynamics. However, it is acceptable for the aims of this work because the simulation is simply used to generate a conformational distribution. The reduced computational complexity of the in vacuo simulation freed up computer time to perform the computationally intensive pK_a calculations.

It should be noted that MD simulations fail to accurately represent true conformational variability. For example, MD simulations tend to have difficulty sufficiently sampling rotamer space,[40] which is why MD simulations sometimes fail to reproduce all NMR side-chain order parameters.[41] Nevertheless, the results presented here reveal clear trends within the conformational dependence of the method and represent a first step toward a better understanding of how conformational variability affects calculated pK_a values. Work is currently underway in extending this investigation to a broader range of protein flexibility as probed by NMR conformational ensembles and various crystallographic isoforms.

Side-chain accessible surface areas (ASAs) are calculated using WhatIf.[42] All ASA values are calculated using a contact surface procedure, meaning WhatIf identifies the molecular surface that a spherical probe (representing a water molecule) can come into contact with. Reported ASA values are for a single exemplar conformer; however, all qualitative conclusions based on that exemplar are robust to the slight structural perturbations. The side-chain structural variability of the titratable residues is calculated as the RMSD of a *target* atom representative of the charge location. RMSDs for each conformer are calculated relative to the average position within the ensemble. Charged residue target atoms are defined by Livesay et al.[35] In the cases of lysine, tyrosine, cysteine, and the N terminus, the target atoms are simply the charged atoms NZ, OH, SG, and N, respectively. In the cases of aspartic acid, glutamic acid, C-terminus histidine, and arginine, the target atoms are CG, CD, C, CE2, and CZ, respectively, which are all central to the multiple partially charged atoms.

Acknowledgment. This research project began as a class exercise in CHM 416 (Macromolecular Modeling) at California State Polytechnic University, Pomona. The class was taught by D.R.L., and the students were J.K., H.C., J.G., P.K., M.P.M., S.P., and D.Y. E.C. is supported by a Howard Hughes Medical Institute undergraduate fellowship. D.R.L. thanks Marty Scholtz and C. Nick Pace for insightful discussions concerning RNase Sa. The reviewers are also thanked for helpful comments.

References

(1) Harris, T. K.; Turner, G. J. Structural Basis of Perturbed PKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life* **2002**, *53*, 85−98.

(2) Kursula, I.; Partanen, S.; Lambeir, A. M.; Antonov, D. M.; Augustyns, K.; Wierenga, R. K. Structural Determinants for Ligand Binding and Catalysis of Triosephosphate Isomerase. *Eur. J. Biochem.* **2001**, *268*, 5189−5196.

(3) Livesay, D. R.; La, D. The Evolutionary Origins and Catalytic Importance of Conserved Electrostatic Networks Within TIM-Barrel Proteins. *Protein Sci.* **2005**, *14*, 1158−1170.

(4) Ha, N. C.; Kim, M. S.; Lee, W.; Choi, K. Y.; Oh, B. H. Detection of Large PKa Perturbations of an Inhibitor and a Catalytic Group at an Enzyme Active Site, a Mechanistic Basis for Catalytic Power of Many Enzymes. *J. Biol. Chem.* **2000**, *275*, 41100−41106.

(5) Rozovsky, S.; Jogl, G.; Tong, L.; McDermott, A. E. Solution-State NMR Investigations of Triosephosphate Isomerase Active Site Loop Motion: Ligand Release in Relation to Active Site Loop Dynamics. *J. Mol. Biol.* **2001**, *310*, 271−280.

(6) Rozovsky, S.; McDermott, A. E. The Time Scale of the Catalytic Loop Motion in Triosephosphate Isomerase. *J. Mol. Biol.* **2001**, *310*, 259−270.

(7) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson−Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15*, 377−392.

(8) Alexov, E. G.; Gunner, M. R. Incorporating Protein Conformational Flexibility into the Calculation of PH-Dependent Protein Properties. *Biophys. J.* **1997**, *72*, 2075−2093.

(9) Alexov, E. G.; Gunner, M. R. Calculated Protein and Proton Motions Coupled to Electron Transfer: Electron-Transfer From QA- to QB in Bacterial Photosynthetic Reaction Centers. *Biochemistry* **1999**, *38*, 8253−8270.

(10) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of PH-Dependent Properties of Proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.

(11) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. The Determinants of PKas in Proteins. *Biochemistry* **1996**, *35*, 7819−7833.

(12) Gilson, M. K. Multiple-Site Titration and Molecular Modeling: Two Rapid Methods for Computing Energies and Forces for Ionizable Groups in Proteins. *Proteins* **1993**, *15*, 266−282.

(13) Nielsen, J. E.; Vriend, G. Optimizing the Hydrogen-Bond Network in Poisson−Boltzmann Equation-Based PK(a) Calculations. *Proteins* **2001**, *43*, 403−412.

(14) van Vlijmen, H. W.; Schaefer, M.; Karplus, M. Improving the Accuracy of Protein PKa Calculations: Conformational Averaging Versus the Average Structure. *Proteins* **1998**, *33*, 145−158.

(15) Zhou, H. X.; Vijayakumar, M. Modeling of Protein Conformational Fluctuations in PKa Predictions. *J. Mol. Biol.* **1997**, *267*, 1002−1011.

(16) Tanford, C.; Roxby, R. Interpretation of Protein Titration Curves. Application to Lysozyme. *Biochemistry* **1972**, *11*, 2192−2198.

(17) Nielsen, J. E.; McCammon, J. A. On the Evaluation and Optimization of Protein X-Ray Structures for PKa Calculations. *Protein Sci.* **2003**, *12*, 313−326.

(18) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Lutty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Gagheri, B.; Scott, L. R.; McCammon, J. A. Electrostatics and Diffusion of Molecules in Solution, Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **1995**, *91*, 57−95.

(19) Gibas, C. J.; Subramaniam, S. Explicit Solvent Models in Protein PKa Calculations. *Biophys. J.* **1996**, *71*, 138−147.

(20) Nielsen, J. E.; Andersen, K. V.; Honig, B.; Hooft, R. W.; Klebe, G.; Vriend, G.; Wade, R. C. Improving Macromolecular Electrostatics Calculations. *Protein Eng.* **1999**, *12*, 657−662.

(21) Kumar, S.; Nussinov, R. Fluctuations in Ion Pairs and Their Stabilities in Proteins. *Proteins* **2001**, *43*, 433−454.

(22) Laurents, D.; Perez-Canadillas, J. M.; Santoro, J.; Rico, M.; Schell, D.; Pace, C. N.; Bruix, M. Solution Structure and Dynamics of Ribonuclease Sa. *Proteins* **2001**, *44*, 200−211.

(23) Laurents, D. V.; Perez-Canadillas, J. M.; Santoro, J.; Rico, M.; Schell, D.; Hebert, E. J.; Pace, C. N.; Bruix, M. Sequential Assignment and Solution Secondary Structure of Doubly Labelled Ribonuclease Sa. *J. Biomol. NMR* **1999**, *14*, 89−90.

(24) Laurents, D. V.; Scholtz, J. M.; Rico, M.; Pace, C. N.; Bruix, M. Ribonuclease Sa Conformational Stability Studied by NMR-Monitored Hydrogen Exchange. *Biochemistry* **2005**, *44*, 7644−7655.

(25) Laurents, D. V.; Huyghues-Despointes, B. M.; Bruix, M.; Thurlkill, R. L.; Schell, D.; Newsom, S.; Grimsley, G. R.; Shaw, K. L.; Trevino, S.; Rico, M.; Briggs, J. M.; Antosiewicz, J. M.; Scholtz, J. M.; Pace, C. N. Charge−Charge Interactions Are Key Determinants of the PK Values of Ionizable Groups in Ribonuclease Sa (PI=3.5) and a Basic Variant (PI=10.2). *J. Mol. Biol.* **2003**, *325*, 1077−1092.

(26) Sevcik, J.; Dauter, Z.; Lamzin, V. S.; Wilson, K. S. Ribonuclease From Streptomyces Aureofaciens at Atomic Resolution. *Acta Crystallogr., Sect. D* **1996**, *52*, 327−344.

(27) Brooks, R. B.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM, A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(28) Livesay, D. R.; Linthicum, S.; Subramaniam, S. PH Dependence of Antibody-Hapten Association. *Mol. Immunol*. **1999**, *36*, 397−410.

(29) Gibas, C. J.; Subramaniam, S.; McCammon, J. A.; Braden, B. C.; Poljak, R. J. PH Dependence of Antibody/Lysozyme Complexation. *Biochemistry* **1997**, *36*, 15599−15614.

(30) Gibas, C. J.; Jambeck, P.; Subramaniam, S. Continuum Electrostatic Methods Applied to PH-Dependent Properties of Antibody−Antigen Association. *Methods* **2000**, *20*, 292−309.

(31) Livesay, D. R.; Subramaniam, S. Conserved Sequence and Structure Association Motifs in Antibody−Protein and Antibody−Hapten Complexes. *Protein Eng., Des. Sel.* **2004**, *17*, 463−472.

(32) Torrez, M.; Schultehenrich, M.; Livesay, D. R. Conferring Thermostability to Mesophilic Proteins Through Optimized Electrostatic Surfaces. *Biophys. J.* **2003**, *85*, 2845−2853.

(33) Zhou, H. X.; Dong, F. Electrostatic Contributions to the Stability of a Thermophilic Cold Shock Protein. *Biophys. J.* **2003**, *84*, 2216−2222.

(34) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Lutty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Gagheri, B.; Scott, L. R.; McCammon, J. A. Electrostatics and Diffusion of Molecules in Solution, Simulations With the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **1995**, *91*, 57−95.

(35) Livesay, D. R.; Jambeck, P.; Rojnuckarin, A.; Subramaniam, S. Conservation of Electrostatic Properties within Enzyme Families and Superfamilies. *Biochemistry* **2003**, *42*, 3464−3473.

(36) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Function for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(37) Sharp, K. A.; Honig, B. Electrostatic Interactions in Macromolecules: Theory and Applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.

(38) Davenport, R. C.; Bash, P. A.; Seaton, B. A.; Karplus, M.; Petsko, G. A.; Ringe, D. Structure of the Triosephosphate Isomerase−Phosphoglycolohydroxamate Complex: An Analogue of the Intermediate on the Reaction Pathway. *Biochemistry* **1991**, *30*, 5821−5826.

(39) Harata, K.; Abe, Y.; Muraki, M. Full-Matrix Least-Squares Refinement of Lysozymes and Analysis of Anisotropic Thermal Motion. *Proteins* **1998**, *30*, 232−243.

(40) Berndt, K. D.; Guntert, P.; Wuthrich, K. Conformational Sampling by NMR Solution Structures Calculated with the Program DIANA Evaluated by Comparison with Long-Time Molecular Dynamics Calculations in Explicit Water. *Proteins* **1996**, *24*, 304−313.

(41) Philippopoulos, M.; Lim, C. Exploring the Dynamic Information Content of a Protein NMR Structure: Comparison of a Molecular Dynamics Simulation With the NMR and X-Ray Structures of *Escherichia coli* Ribonuclease HI. *Proteins* **1999**, *36*, 87−110.

(42) Vriend, G. WHAT IF: A Molecular Modeling and Drug Design Program. *J. Mol. Graphics* **1990**, *8*, 52−6, 29.

CT600066Z