Biophysical Journal, Volume 98

Supporting Material

Statistics and physical origins of pK and ionization state changes upon protein-ligand binding

Boris Aguilar, Ramu Anandakrishnan, Jory Z. Ruscio, and Alexey V. Onufriev

Supporting Material

Brief summary of the methodology for pK estimates

The standard continuum electrostatics methodology (1) was used to compute pK values; the physical conditions such as dielectric constants used are given in the "Methods" section of the main text. Within our computational model, the difference between a residue's pK and the pK of the corresponding model compound in solution is determined by the combined effect of two distinct contributions to the total electrostatic (free) energy change. First is the "Born term" or desolvation penalty, which always penalizes burial of a charge inside a low dielectric medium. Second is the "background term", which represents the electrostatic interactions of the residue in question with all other fixed charges in the molecule not belonging to any ionizable residues. These energy terms, as well as the matrix of site–site interactions W_{ij} are estimated through a sequence of finite-difference Poisson-Boltzmann calculations in which residues in the protein and their corresponding model compounds have their charge distributions set to those of the protonated or deprotonated form, and suitable energy differences are taken. In the finite difference lattices, several levels of focusing are used. In the coarsest level the bounding box is set to twice the molecule's maximum extent and the grid points are spaced 2 Å apart. The finest lattice is "focused" on the region of interest, and the grid points are 0.5 Å apart. The probe radius for defining the molecular surface, which is used as the boundary between the interior and exterior dielectric regions, is set to 1.4 Å.
The electrostatics calculation outlined above provides (free) energies of each of the 2^N protonation micro-states in the system, where N is the number of ionizable residues. Based on these free energies, H++ uses Boltzmann averages to compute the probability of protonation of each residue at different pH values, which define the titration curves from which pK values are calculated as mid-points.

Validation

In what follows, we (i) estimate the average systematic error of the computed ΔpK values by comparison with the available experimental numbers, and (ii) provide additional material used in the estimate of random error.

Systematic error of computed $|\Delta\text{p}K|$ — comparison with experiment

We compared the calculated pK changes with the corresponding experimental values for 13 cases where experimental data for pK changes upon binding are available, see Table S1. In most cases, the direction of the computed pK change (the sign) is in agreement with experiment. The magnitude of pK changes for individual residues is also in reasonable agreement with the experiment. Most importantly, the absolute value of the average computed pK change is in quantitative agreement with the corresponding experimental value. Specifically, the experimentally determined ΔpKs shown in Table S1 have a root

mean square (RMS) of 2.91 pK units. The RMS of the computationally determined ΔpKs is 3.23. Thus, our methodology is able to predict, on average, substantial pK changes ($|\Delta \text{p}K| > 1.0$) upon ligand binding.

Table S1: Comparison between the computed and experimental pK changes, ΔpK, upon ligand binding. Each ΔpK value was computed as the difference between the pK values of the residue in the complex and in the unligated protein. The "overall" procedure was used, see "Methods" in the main text. In some cases two ionizable residues are so strongly coupled that their individual pKs may lose their conventional meaning (2)[†]. Such residues are listed together, separated by a "/" in the table. In the case of HIV protease dimer, ASP25 exists in each of its two chains; the two residues are distinguished by labeling one as ASP25 and the other as ASP25'. Experimental literature reference is given in ( ), next to each data point.

Site	Protein PDB		pK change	
	In absence of ligands	In complex with a ligand	Experiment	Computation
ASP25	3HVP	1HPX	>+2.2(3, 4)	+16.3
ASP25'	3HVP	1HPX	<-1.5(3, 4)	-0.5
ASP34/214	1PFZ	1SME	+0.1(5)	+5.3
HIS164	1PFZ	1SME	+1.5(5)	+0.5
APS33/231	1LYW	1LYB	> 0(5)	+7.8
HIS77	1LYW	1LYB	> 0(5)	+2.1
GLU260	1LYW	1LYB	> 0(5)	-2.6
ASP25/25'	1QBS	1QBS	+2.19(6)	+4.1
ASP25/25'	1HHP	1QBS	+2.19(6)	+3.7
HIS57	6GCH	6GCH	+3.3(7)	-0.4
GLU172	1BVV	1BVV	-2.5(8)	-1.3
ASP35/GLU172	1C5I	1C5I	< 0.2(8)	+0.2
HIS235	2YAS	2YAS	+5.5(9)	+3.5

†The proper single pK of such pairs is obtained from a combined titration curve, which is defined as the average of the computationally determined protonation probability for the two residues, at different pH values.

Complementary material for the random error estimation

Dataset for the random error estimate

Table S2: Collection of 7 proteins used in the estimation of the false positive level in the percentages of ionizable residues that exhibit substantial pK change in protein-ligand binding. For each of the proteins listed, several X-ray structures of its unligated form are available. Reported experimental conditions, as well as specific crystal symmetry groups, differ between many of these structures, suggesting that the contribution of the corresponding factors to the "structural noise" is taken into account by our estimate.

Protein Name	PDB IDs of unligated structures
FAB E8	1QBM, 1QBL
Subtilisin	1SUP, 1SBT, 1S01
Cyt C peroxidase	1CCP, 1DCC, 1KOK, 1SBM, 2CYP, 1Z53, 1ZBY, 1ZBZ, 1CPD
Ferredoxin reductase	1GJR, 1QUE, 1QUF
p67 Phox	1HH8, 1WM5
alpha-amylase	1PIG, 1HX0, 1JFH, 1OSE, 1PIF, 1VAH, 1WO2
Ras GTPase	6Q21, 5P21, 4Q21, 821P, 1GNR, 1JAH

t-test analysis of the structural "noise" (random error)

We use a t-test to compare the mean (average) of the number of substantial pK changes due to structural "noise" ("noise" sample) to the mean of the number of substantial pK changes due to ligand binding ("signal" samples). An analysis presented in the main text shows that it is only the significance of pK changes outside the binding interface region that may be in question, and so we apply the t-test only to residues out of the interface region (more than 6 Å away from the ligand). Four data samples are constructed for the test: three "signal" samples, and the "noise" sample.

We set up three "signal" samples, one for protein-protein complexes, one for protein-small molecule, and one for protein-nucleic acid complexes. Each "signal" sample consists of 20 data points which represent the percentage of ionizable residues out of the interface that exhibit substantial and biologically relevant pK changes due to ligand binding.

The "noise" sample contains 80 data points, one for each pair of unligated structures presented in Table S2. Each data point represents the percentage of residues located outside the interface region that exhibit substantial difference between the pK values of the pair of unligated structures. As before, we considered only the pK values within the biologically relevant pH range from 4-8. It is worth mentioning that the proteins in Table S2 were selected from the protein-protein data set, so the set of ionizable residues inside and outside the interface region of the proteins used in the "noise" sample is identical to the set of ionizable residues in the corresponding proteins of the "signal"

sample. We performed a t-test of the null hypothesis that data in the "noise" and the "signal" samples have equal means, against the alternative hypothesis that the means are not equal. The results of the t-test of the null hypothesis for a probability of 0.95 are presented in table S3. Thus, we conclude that the computed substantial pK changes outside of the binding interface region are statistically significant in the protein-protein and protein-nucleic acid complexes. In the case of protein-small molecule complexes we do not have enough statistics to confirm or reject the possibility of substantial changes outside the binding interface region.

Table S3: Results of the t-test of the null hypothesis that data in the "noise" sample and the "signal" samples have equal means, with a probability of 0.95. The mean of the noise sample is 4.2%

Complex type	Mean "Signal"	p-value	t-test conclusion
Protein-small molecule	4.6%	0.56	Null hypothesis NOT rejected
Protein-protein	6.7%	< 0.05	Null hypothesis rejected
Protein-nucleic acid	10.0%	< 0.05	Null hypothesis rejected

Effect of re-arrangement of hydrogen positions in response to "electrostatic" changes

To further dissect the various physical contributions to substantial pK changes seen in figure 2 of the main text, we have investigated the effect of re-arrangement of hydrogen positions in response to ligand binding that excluded any re-arrangements of heavy atoms due to "induced-fit" (that is assuming that heavy atoms did not move). Clearly, some re-arrangement of hydrogen positions and possibly of hydrogen bonding networks may occur purely due to solvation/de-solvation of the binding interface and changes in the charge state of its immediate surroundings caused by ligand binding. In other words, these are proton position changes caused by the "electrostatic only" effects introduced earlier. To single out a possible effect of such changes upon pK of ionizable residues in the receptor, we have introduced the fourth procedure for computing ΔpK. This procedure is shown in Figure S1. Basically, we compare the pK values of the ionizable sites of the unligated proteins resulting from the "electrostatics only" and "structural changes only" procedures, see the "Methodology" section of the main text for details. The only difference between these two unligated structures is the positions of their hydrogen atoms that have responded to the binding of the ligand (have been computationally re-optimized in the presence of the ligand) .

The black solid bars of figure S2 represent the pK changes due to hydrogen optimization. Clearly, the effect of this specific re-arranegemnt of hydrogen atoms is very small, thus its contribution to the overall effect on substantial pK changes can be safely neglected. Moreover, these results suggest that the substantial pK changes far away from the interface that we have seen are only due to the re-arrangement of heavy atoms

caused by "induced fit". Of course, such conformational changes will cause movement of hydrogens, but the causal origin of these movements is the "induced fit" effects already analyzed (blue bars in figure S2), rather than alteration of the "electrostatic" state of the binding interface propagated out through alterations in H-bonding network.

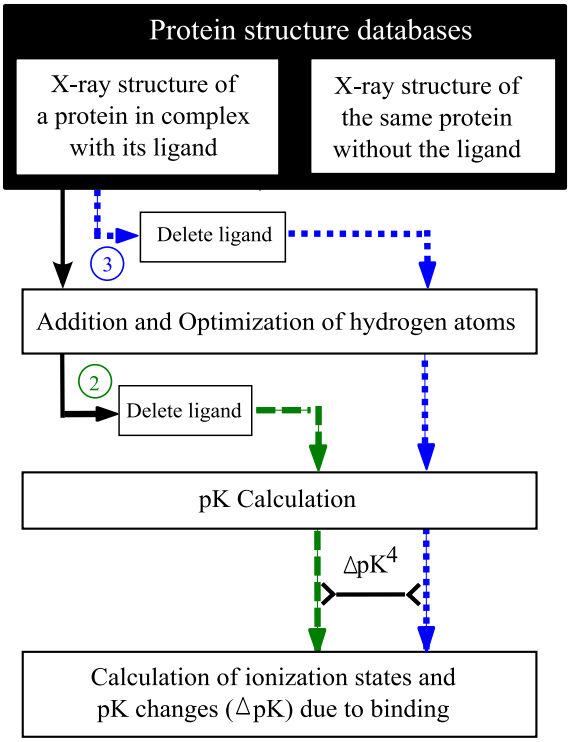

Figure S1: Flowchart of the computational methodology used to estimate the effect of hydrogen optimization.

Analysis of local RMS deviation in regions far from the binding interface

In order to further explore the nature of the structural changes responsible for the substantial pK changes in regions far (>12 Å) from the interface, we calculate the "local RMS deviation" for each ionizable residue of our protein-protein data set. We define "local RMS deviation" as the RMS between the heavy atoms of the complexed and unligated structures of one particular residue and its neighbors (residues within 5 Å).

We compute "local RMS deviation" for ionizable residues far from the interface that exhibit substantial pK changes, and for ionizable residues far from the interface with no substantial pK change, then we compare both samples. Figure S3 shows the histograms

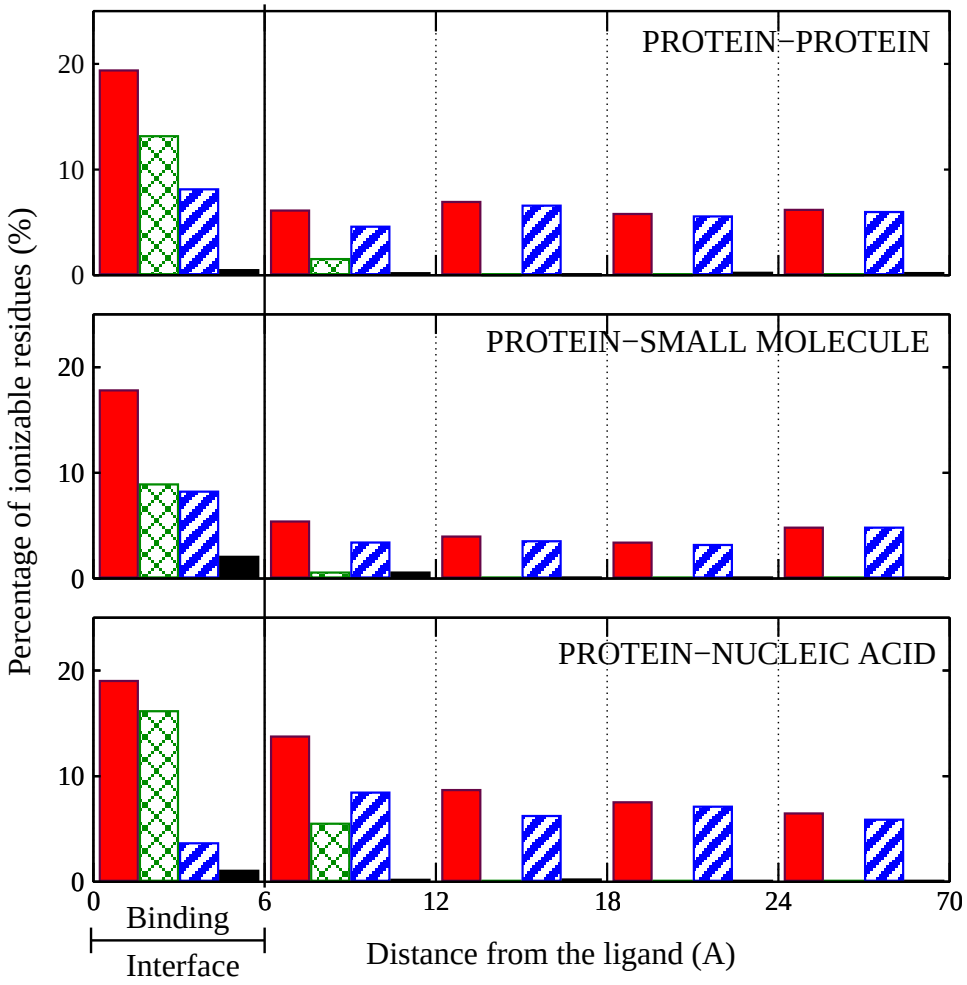

Figure S2: Distance distribution of ionizable residues with substantial pK change in the biologically relevant pH range. Color coding consistent with figure 2 of the main text. Red bars: "overall" procedure, green bars: "electrostatics only" procedure, blue bars: "structural changes only" procedure, and black bars: effect of re-arrangement of hydrogen positions in response to ligand binding, assuming no movement of heavy atoms.

of the samples described above. Both histograms look somewhat similar, however, the average of "local RMS deviation" of residues that exhibit substantial pK changes is larger than those without substantial pK change, 1.06 and 0.92 respectively. This distinction is statistically significant: a t-test rejected the hypothesis ($p<0.05$) that both samples were identical.

These results suggest that small structural changes can produce substantial pK changes. Moreover, the similarity, broadness, and the absence of distinctive features in Figure S3 for both samples suggest that there is no single dominant structural mechanism responsible for pK changes far from the binding interface.

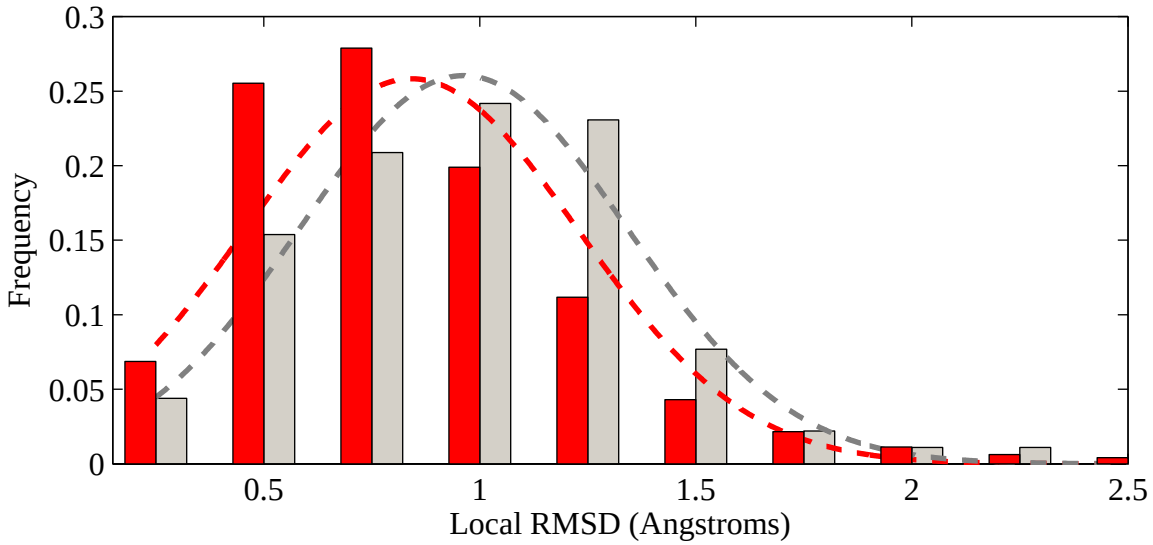

Figure S3: Local Root- Mean-Square Deviation (RMSD) of positions of ionizable residues that exhibit substantial pK changes (Grey) versus ionizable residues with no substantial pK changes (Red). Here we consider only the residues of the protein-protein dataset located in regions far (>12 Å) from the binding interface.

Examples of types of conformational changes that correlate with substantial pK changes

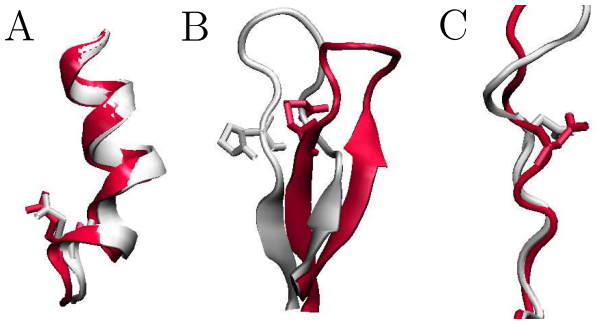

Figure S4: Examples of types of structural changes between unligated (Red) and complexed (light Grey) forms that induce substantial pK changes in some residues (thick lines). (A) An entire helix performs a small tilt movement that affect the pK value of GLU-720 of PDB 6Q21. (B) The residue ARG-310 of PDB 1E1N experiences a substantial pK shift due to random-coil to beta-sheet transition. (C) A ionizable residue (ASP-982 of PDB 6Q21) with substantial pK change located in a flexible random coil region experiences a large backbone movement due to complex formation.

Protein-ligand datasets

Table S4: Protein-protein dataset. We also include the total number of ionizable residues and the number of residues with substantial (computed) pKchange in the biologically relevant range ($|\Delta \mathrm{p}K| > 1$), inside and outside the binding interface region.

PDB ID of complexes	Unligated proteins		# of residues Inside interface		# of residues Outside interface	
	PDB ID 1	PDB ID 2	Total	$\|\Delta \mathrm{p}K\| > 1$	Total	$\|\Delta \mathrm{p}K\| > 1$
1VFB	1VFA	8LYZ	20	2	59	4
1WEJ	1QBL	1HRC	23	1	116	5
2SIC	1SUP	3SSI	15	3	49	5
1HE1	1MH1	1HE9	18	1	67	7
1FQJ	1TND	1FQI	27	5	125	1
1GHQ	1C3D	1LY2	10	1	96	6
2SNI	1UBN	2CI2	15	0	50	1
1CGI	2CGA	1HPT	19	2	37	5
1E6E	1E1N	1CJE	25	7	141	8
2PCC	1CCP	1YCC	18	4	122	4
7CEI	1UNK	1M08	18	7	66	7
1EWY	1GJR	1CZP	24	7	111	13
1E96	1MH1	1HH8	13	3	110	4
2MTA	2BBK	2RAC	16	3	157	11
1A2K	1QG4	1OUN	15	3	112	5
1BVN	1PIG	1HOE	26	6	121	6
1DFJ	9RSA	2BNH	44	12	124	9
1MLC	1MLB	3LZT	13	1	113	7
1TMQ	1JAE	1B1U	27	7	115	6
1WQ1	6Q21	1WER	32	6	125	13

Table S5: Protein-small molecule dataset. We also include the total number of ionizable residues and the number of residues with substantial (computed) pK change in the biologically relevant range ($|\Delta \mathrm{p}K| > 1$), inside and outside the binding interface region.

PDB ID of The complex	PDB ID of the Unligated protein	# of residues Inside interface		# of residues Outside interface	
		Total	$\|\Delta \mathrm{p}K\| > 1$	Total	$\|\Delta \mathrm{p}K\| > 1$
1AWQ	2CPL	4	1	45	2
1JET	1RKM	14	4	153	7
1AYC	1AYD	10	0	23	3
1YE3	8ADH	2	0	103	4
1NDV	1VFL	8	0	121	15
1CBX	5CPA	10	2	68	0
2OY2	2OY4	16	5	72	1
1G46	1FQL	4	1	78	5
1PGP	2PGD	7	2	135	5
1PSO	1PSN	6	2	49	0
1RBP	1BRQ	7	1	51	3
1XIG	1XIB	12	5	117	1
3GBP	1GCG	10	1	79	3
2IFB	1IFB	8	0	39	1
5TIM	2V5L	1	0	59	1
1TYD	1TYC	9	1	97	4
1BE8	1SCD	2	0	43	2
1WOO	1WOS	10	1	102	3
2IZJ	2RTA	3	0	21	3
1QGF	3EST	3	0	40	1

Table S6: Protein-nucleic acid dataset. We also include the total number of ionizable residues and the number of residues with substantial (computed) pK change in the biologically relevant range ($|\Delta \mathrm{p}K| > 1$), inside and outside the binding interface region.

PDB ID of The complex	PDB ID of the Unligated protein	# of residues Inside interface		# of residues Outside interface	
		Total	$\|\Delta \mathrm{p}K\| > 1$	Total	$\|\Delta \mathrm{p}K\| > 1$
3HDD	1ENH	6	2	12	2
1J59	1G6N	25	5	98	25
1M5O	1OIA	24	3	27	2
2E1C	2CYY	8	2	43	2
1MJM	1MJK	10	1	74	9
2AC0	2OCJ	40	3	224	16
1EMH	1AKZ	9	3	61	3
1RCN	1C9X	9	3	23	1
1RVA	1RVE	33	8	149	13
2PVI	1K0Z	29	6	77	8
1A74	1EVX	26	5	60	8
1G9Z	2O7M	40	8	54	11
1QUM	1QTW	17	3	71	3
1G59	1GLN	40	13	133	14
2HW8	1AD2	16	3	56	3
1SDS	1RA4	6	0	33	5
1M8W	1IB2	20	1	87	6
1ZBL	1ZBF	8	2	36	4
2F8K	2D3D	8	1	23	2
2A8V	1A62	10	1	26	1

References

1. Bashford, D., and M. Karplus. 1990. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry.* 29:10219–10225.

2. Onufriev, A., D. A. Case, and G. M. Ullmann. 2001. A novel view of pH titration in biomolecules. *Biochemistry.* 40:3413–3419.

3. Smith, R., I. M. Brereton, R. Y. Chai, and S. B. Kent. 1996. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Mol. Biol.* 3:946–950.

4. Wang, Y. X., D. I. Freedberg, T. Yamazaki, P. T. Wingfield, S. J. Stahl, J. D. Kaufman, Y. Kiso, and D. A. Torchia. 1996. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry.* 35:9945–9950.

5. Xie, D., S. Gulnik, L. Collins, E. Gustchina, L. Suvorov, and J. W. Erickson. 1997. Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: Direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry.* 36:16166–16172.

6. Yamazaki, T., L. K. Nicholson, P. Wingfield, S. J. Stahl, J. D. Kaufman, C. J. Eyermann, N. C. Hodge, P. Y. Lam, and D. A. a. Torchia. 1994. NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.* 116:10791–10792.

7. Cassidy, C. S., J. Lin, and P. A. Frey. 1997. A new concept for the mechanism of action of chymotrypsin: The role of the low-barrier hydrogen bond. *Biochemistry.* 36:4576–4584.

8. Joshi, M. D., G. Sidhu, J. E. Nielsen, G. D. Brayer, S. G. Withers, and L. P. Mcintosh. 2001. Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. *Biochemistry.* 40:10115–10139.

9. Stranzl, G. R., K. Gruber, G. Steinkellner, K. Zangger, H. Schwab, and C. Kratky. 2004. Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from hevea brasiliensis explains a large pK_a shift of the catalytic base induced by the reaction intermediate. *J. Biol. Chem.* 279:3699–3707.