Long-Range Electrostatic Contributions to Protein−Ligand Binding Estimated Using Protein Charge Ladders, Affinity Capillary Electrophoresis, and Continuum Electrostatic Theory

Justin A. Caravella,[†] Jeffrey D. Carbeck,[‡,§] David C. Duffy,[‡,⊥] George M. Whitesides,*[,‡] and Bruce Tidor*[,†]

Contribution from the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received December 7, 1998

Abstract: Affinity capillary electrophoresis and protein charge ladders are used together to measure the contributions of long-range electrostatic interactions to binding of substituted benzene sulfonamide inhibitors to derivatives of human carbonic anhydrase II. The results are analyzed by continuum electrostatic calculations, which afford a detailed analysis of interactions of individual members of a population from a charge ladder. A Monte Carlo simulation of the experimental data using calculated contributions of individual lysine side chains to inhibitor binding shows that a large number of different patterns of acetylation are consistent with the experimental results. The calculations predict significant differences in the contributions of some lysines to ΔG and simulations suggest that experimental resolution must be enhanced to be able to measure such differences.

Introduction

Electrostatic interactions affect the affinity, specificity, and catalytic properties of biomolecules. Analysis of electrostatic effects is complicated by their long-range nature and by the observation that their net effect often results from opposing contributions, such as unfavorable desolvation penalties that are at least partially offset by favorable intermolecular interactions on binding. One particularly striking long-range electrostatic effect involves the Met repressor binding to its DNA operator. In this system, the binding affinity increases 1000-fold upon binding a positively charged cofactor; this increase has been attributed almost entirely to a long-range electrostatic attraction between the charged cofactor and the DNA.[1,2] In another study, theoretical methods were used by Allewell and co-workers to suggest that linked networks of long-range electrostatic interactions could be important in the conformational changes of aspartate transcarbamylase.[3]

Experiments devised to evaluate electrostatic interactions include measuring pK_a shifts of ionizable groups due to modification of other charged groups[4] and the related method of double-mutant cycles.[5−8] These interactions can also be studied with theoretical methods, although theoretical estimation of pK_a shifts has had only modest success.[9−12] Among other difficulties, conformational relaxation accompanying mutation or titration limit agreement between experiment and theory.[13−16]

The combination of protein charge ladders and affinity capillary electrophoresis (ACE) is useful for estimating the role of electrostatic interactions in binding affinity.[17,18] The method overcomes some of the difficulties of other approaches and offers some unique advantages. In the present study, human carbonic anhydrase II (HCA II) was modified to produce a protein charge ladder—a mixture of protein derivatives in which various numbers of lysine ϵ-amino groups are acetylated and each acetylation alters the nominal protein charge by one. The interactions between the components of this mixture of proteins and three differently charged benzene sulfonamide inhibitors of HCA II were measured experimentally using ACE. The use of differently modified proteins in combination with measure-

* Authors to whom correspondence should be addressed.
† Massachusetts Institute of Technology.
‡ Harvard University.
§ Current address: Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544.
⊥ Current address: Gamera Bioscience, 200 Boston Avenue, Medford, Massachusetts 02155.

(1) Somers, W. S.; Rafferty, J. B.; Phillips, K.; Strathdee, S.; He, Y.-Y.; McNally, T.; Manfield, I.; Navratil, O.; Old, I. G.; Saint-Girons, I.; Stockley, P. G.; Phillips, S. E. V. *Ann. N.Y. Acad. Sci.* **1994**, *726*, 105.
(2) Phillips, K.; Phillips, S. E. V. *Structure* **1994**, *2*, 309.
(3) Oberoi, H.; Trikha, J.; Yuan, X.; Allewell, N. M. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 300.
(4) Loewenthal, R.; Sancho, J.; Reinikainen, T.; Fersht, A. R. *J. Mol. Biol.* **1993**, *232*, 574.
(5) Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. *Cell* **1984**, *38*, 835.
(6) Horovitz, A. *J. Mol. Biol.* **1987**, *196*, 733.
(7) Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. *Biochemistry* **1990**, *29*, 9343.
(8) Schreiber, G.; Frisch, C.; Fersht, A. R. *J. Mol. Biol.* **1997**, *270*, 111.
(9) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219.
(10) Gilson, M. K. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 266.
(11) You, T. J.; Bashford, D. *Biophys. J.* **1995**, *69*, 1721.
(12) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Biophys. J.* **1995**, *68*, 2233.
(13) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819.
(14) Simonson, T.; Brooks, C. L., III *J. Am. Chem. Soc.* **1996**, *118*, 8452.
(15) Simonson, T. *J. Am. Chem. Soc.* **1998**, *120*, 4875.
(16) Beroza, P.; Case, D. A. *J. Phys. Chem.* **1996**, *100*, 20156.
(17) Gao, J.; Mammen, M.; Whitesides, G. M. *Science (Washington, D.C.)* **1996**, *272*, 535.
(18) Chu, Y.-H.; Avila, L. Z.; Gao, J.; Whitesides, G. M. *Acc. Chem. Res.* **1995**, *28*, 461.

10.1021/ja984195a CCC: $18.00
Published on Web 04/27/1999

ments on differently modified inhibitors effectively provides a double-mutant cycle, which we use to measure interaction energies involving lysine residues in HCA II and their effect on binding benzene sulfonamide inhibitors. The advantage of this method over making individual mutant proteins is that it gives an interaction free energy that represents an average over many possible "mutants" (proteins with sets of acetylated lysines) without the enormous effort that would be required to make individual mutations of different combinations of lysines. The precision of this measurement of binding energy may be improved because it represents an average over many interactions. One potential disadvantage is that modified proteins with the same net charge are not separated from each other and their individual binding properties cannot be measured.

The experiments are complemented by continuum electrostatic calculations that compute individual contributions of each lysine to the binding affinity. The calculations are based on the crystal structure of HCA II bound to a benzene sulfonamide inhibitor.[19] Contributions of individual lysines must be appropriately averaged to compare to the experimental (population) results. Through the use of a variety of different populations and the simulation of an ACE experiment, we demonstrate that the computed results are consistent with a variety of different acetylation patterns.

Experimental Approach

The Formation and Analysis of Charge Ladders. The treatment of HCA II with acetic anhydride results in the formation of a set of derivatives that differ in their number and distribution of acetylated Lys ϵ-amino groups and, therefore, their net charge; the coefficient of friction is approximately unaffected by the modification. The change in charge upon acetylation of Lys ϵ-amino groups depends on the value of pH of the electrophoresis buffer (8.4 in these studies) and the pK_a of Lys ($\sim$10.2) and is therefore assumed to be -1.

This mixture of derivatives of proteins is analyzed by capillary electrophoresis (CE), which separates the components of the mixture based on their electrophoretic mobility (directly related to the net charge of the protein, and inversely related to the coefficient of friction of the protein with the solvent). CE groups the distribution of modified proteins according to their net charge, or equivalently, their number of acetylated Lys ϵ-amino groups, into individual peaks or "rungs" of a charge ladder: the rung with the least negative charge is the native HCA II, and is referred to as rung 0; rung 1 is a peak consisting of those proteins that have a single acetylated Lys ϵ-amino group, and is a mixture of up to 24 regioisomeric derivatives of the protein. The set of rungs of a charge ladder appears in CE as a set of peaks whose spacing varies in a regular way with the number of acetylated Lys ϵ-amino groups (bottom plot in Figure 1).

Affinity Capillary Electrophoresis. We measure the binding affinity of the derivatives that make up the charge ladder of the protein for various ligands using affinity capillary electrophoresis (ACE). In this technique the mobilities of the rungs are measured at various concentrations of ligand added to the electrophoresis buffer; the change in mobility with concentration of ligand is assumed to be directly proportional to the fraction of protein that is bound to ligand. Scatchard analysis of the mobility as a function of ligand concentration yields a dissociation constant, K_d, and a standard free energy of binding, ΔG. Figure 1 shows stacked electropherograms of the charge ladder of HCA II in the presence of increasing concentrations of inhibitor 1 in the electrophoresis buffer and illustrates how the mobility of the rungs of the charge ladder are affected by ligand. The ligand is

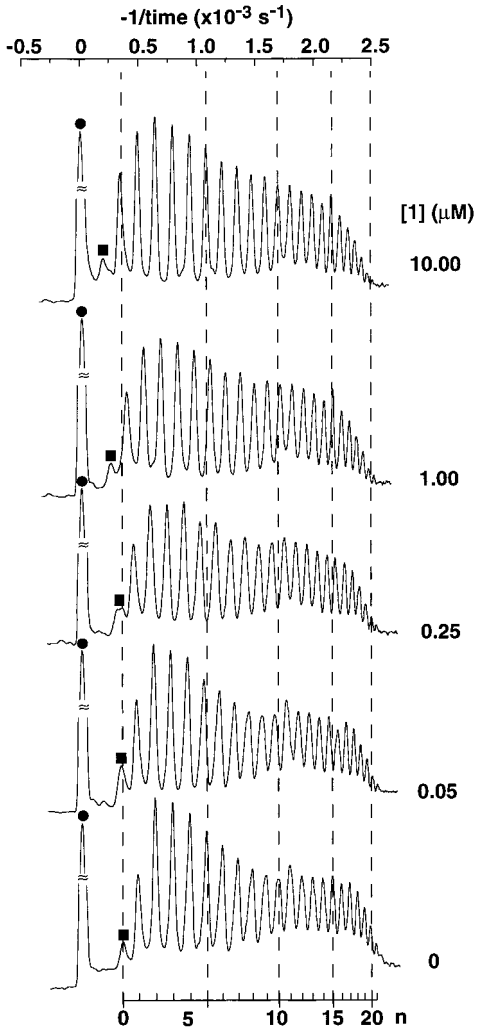


Figure 1. Electropherograms illustrating the change in the electrophoretic mobility of the rungs of the charge ladder of HCA II with increasing concentrations of inhibitor 1. The number of acetylated ϵ-amino groups (n) are estimated below the electropherogram. The peak marked with the circle is a neutral marker that measures electroosmotic flow, and the peak marked with the square is the native protein.

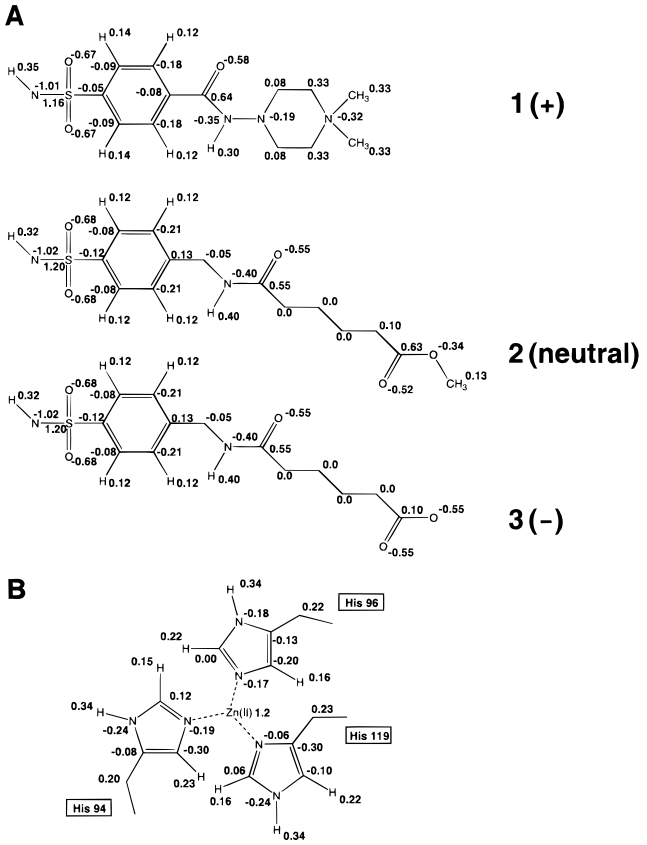
positively charged and the effective charge on each derivative of the protein is changed by an amount that depends on the amount of ligand bound to the protein (i.e., the fraction of time spent in the bound state). At sufficiently high ligand concentrations, the protein is saturated and its apparent charge differs from that of the protein in the unbound state (i.e., in the absence of ligand) by an amount equal to the charge on the ligand. At all concentrations of ligand, the individual rungs of the charge ladder of HCA II are resolved and the amount of ligand bound to the proteins that make up each rung of the charge ladder can be measured in a single set of experiments.

In this study, three substituted benzene sulfonamides are used as ligands for HCA II. The structures of these inhibitors, along with the partial atomic charges used in the calculations, are shown in Scheme 1. The pendant group substituted in the *para* position determines the charge of these inhibitors in solution: inhibitor 1 is positively charged, inhibitor 2 is neutral, and inhibitor 3 is negatively charged.

Theoretical Approach

The structure of HCA II bound to a benzene sulfonamide inhibitor is shown in Figure 2.[19] There are 23 lysines in the

Scheme 1. (A) Structures of the Benzene Sulfonamide Inhibitors Substituted in the Para Position with Neutral and Charged Pendant Groups[a] and (B) Partial Atomic Charges Used on the Zn^{2+} and the Histidine Side Chains in the Active Site of HCA II.



[a] The partial atomic charges used are shown next to the atoms.

crystallographic model, each of which may be acetylated in the experiment. The 24th lysine (Lys-261) is disordered in the crystal structure and is omitted from the analysis. The N-terminus of the unmodified protein is preacetylated and is therefore not reactive here. Figure 2 also shows the Zn^{2+} ion in the binding site, which is coordinated by three histidine residues and the sulfonamide group of the inhibitor.

In the unbound state of HCA II, the Zn^{2+} ion is also coordinated by a water whose pK_a is 7[20] (which becomes hydroxide at the experimental pH of 8.4). Hydroxide is displaced upon binding the ligand, which is believed to be deprotonated in the bound state. The net change in charge at the binding site is simply the charge on the pendant group of the inhibitor, as illustrated in Figure 3. Models for each of the three inhibitor complexes were constructed and used in the analysis of long-range electrostatic effects. Figure 4 is a histogram of the distances of the lysines from the formal positive charge on inhibitor 1. Lys-170 of HCA II is only 10 Å from this nitrogen, while the other 22 lysines are 17-38 Å from it.

The experiment involves acetylation of the various lysines in HCA II. We wish to determine the difference in binding free energy, $\Delta\Delta G$, caused by acetylating a lysine. Figure 3 schematically shows two binding reactions. The first is between a ligand and HCA II with a fully charged lysine, and its binding free energy is ΔG_{b1}. The second is the same process with an acetylated, uncharged lysine, and its binding free energy is ΔG_{b2}. The difference in binding free energy upon acetylation of a lysine ($\Delta G_{b2} - \Delta G_{b1}$) is related to the free energy of charging a lysine in the unbound state (ΔG_{c1}) and that of charging a lysine in the bound state (ΔG_{c2}):

$$\Delta\Delta G = \Delta G_{b2} - \Delta G_{b1} = \Delta G_{c1} - \Delta G_{c2} \qquad (1)$$

To calculate $\Delta\Delta G$ for each lysine, we determine the free energies of charging ΔG_{c1} and ΔG_{c2} using continuum electrostatic theory. The contribution of each lysine to the overall binding affinity should be entirely electrostatic, since all of the lysines are far (>10 Å) from the binding site. Calculations on acetylated lysines show that the contribution of the polar *N*-acetyl group to $\Delta\Delta G$ are negligible (less than 5×10^{-3} kcal/mol; results are not shown). The analysis is therefore simplified to understanding the effect of adding a charge to the lysine, illustrated in Figure 3.

The electrostatic free energy of ligand binding can be divided into three parts: (1) the solvent-screened direct Coulombic interactions between charged and polar groups in the protein and charged and polar groups in the ligand/hydroxide, (2) the change in interactions of the protein and ligand with solvent, and (3) the *intra*molecular interactions of the protein and the ligand, whose magnitude is changed upon binding because of the change in screening by the solvent and ion environment. These three contributions are illustrated in Scheme 2, where they are labeled as direct, solvation, and indirect, respectively.

We computed changes in the three types of contributions due to lysine acetylation. Since we assume that a lysine is the only group whose charge is significantly altered by acetylation, we need only consider how each lysine contributes to the interactions enumerated above. A more detailed analysis would compute changes in the titration behavior of all residues (especially His, Cys, and Tyr, of which there are 12, 1, and 8, respectively) due to lysine acetylation, though these effects are expected to be small due to solvent screening.

Results

Dependence of ΔG on the Number of Acetylations of the Protein. ACE was used to measure ΔG of binding for each rung of the charge ladder to each of the three ligands. The plot of ΔG *versus* rung number (n) is approximately linear for the three inhibitors, as observed in previous, similar experiments using bovine carbonic anhydrase II.[17] The slopes of this plot give the average interaction energy between a lysine and the sulfonamide inhibitor (solid lines in Figure 5).

The slope for positively charged inhibitor 1 is −0.07 kcal/mol per acetylation, and that for negatively charged inhibitor 3 is +0.05 kcal/mol per acetylation, since the net negative charge on the protein is increased by acetylation. The slope for inhibitor 2 (the neutral inhibitor) is −0.01 kcal/mol per acetylation. One method of analyzing these results is through the standard double-mutant cycle argument. The appropriate difference in slopes gives the average interaction energy between a lysine and a charged pendant group: −0.06 kcal/mol for the positively charged inhibitor (difference of slopes 1 and 2) and +0.06 kcal/mol for the negatively charged inhibitor (difference of slopes 3 and 2).

Calculated Interactions of Individual Lysines. The contribution to $\Delta\Delta G$ for each lysine was calculated, as well as the $\Delta\Delta G_{solv}$, $\Delta\Delta G_{dir}$, and $\Delta\Delta G_{indir}$ terms. These contributions to $\Delta\Delta G$ for each lysine depend only on its interactions with solvent, with the ligand at the active site, and with other groups

(19) Boriack, P. A.; Christianson, D. W.; Kingery-Wood, J.; Whitesides, G. M. *J. Med. Chem.* **1995**, *38*, 2286.

(20) Lindskog, S.; Engberg, P.; Forsman, C.; Ibrahim, S. A.; Jonsson, B.-H.; Simonsson, I.; Tibell, L. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 61.

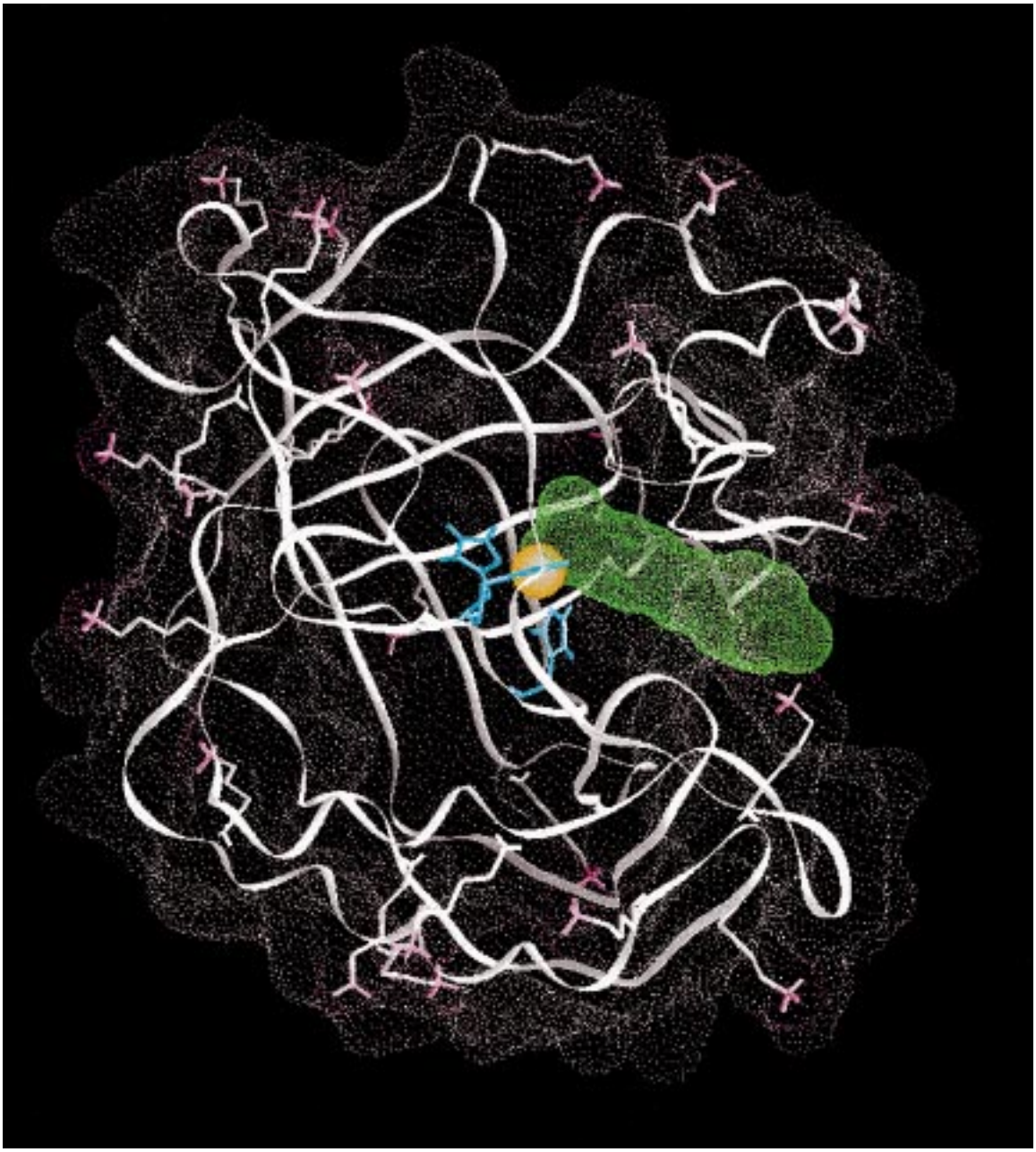


Figure 2. The structure of HCA II. The lysine side chain nitrogens and hydrogens are in pink. The active site zinc ion is shown as a gold sphere, and the side chains of the histidines ligated to it are shown in blue. The surface of the protein is depicted with white dots, and the surface of the bound inhibitor (inhibitor 1) is shown with green dots. All of the lysines are surface accessible, as indicated by the pink dots where the lysine amino groups contact the surface.

in HCA II. Both the ligands and the groups whose dielectric environment change are a long distance from most lysines (at least 17 Å, see Figure 4). As a result, the calculated values of $\Delta\Delta G$ are relatively small—on the order of hundredths of a kcal/mol. However, the long distances over which the interactions occur also means that the calculations are more precise. For example, conformational flexibility of lysine side chains should not substantially alter such long-range interactions. Interactions with groups that are near the lysines generally do not change upon ligand binding, and therefore they also should make no contribution to $\Delta\Delta G$. Small values for $\Delta\Delta G_{solv}$ (less than 4×10^{-4} kcal/mol for all lysines, except Lys-170) were expected because the region of the protein that is desolvated on ligand binding is far from all of the lysines other than Lys-170 (maximum $\Delta\Delta G_{solv}$ of 0.02 kcal/mol).

The average values of the total calculated $\Delta\Delta G$ are −0.07, 0.00, and +0.07 for inhibitors 1, 2, and 3, respectively. The dashed lines in Figure 5 are plotted with these slopes. The three lines originate from the points for the zeroth rung of the charge ladder. Figure 6 shows the distributions of calculated values for $\Delta\Delta G_{dir}$, $\Delta\Delta G_{indir}$, and the total $\Delta\Delta G$ for each of the three inhibitors. Lys-170 has an especially large value of $\Delta\Delta G$ in the presence of inhibitors 1 and 3. Larger values of $\Delta\Delta G$ are expected for Lys-170 because it is rather close to the formal charge on these inhibitors. Its calculated contribution to $\Delta\Delta G$ is 0.68 kcal/mol for inhibitor 1 and −0.54 kcal/mol for inhibitor 3.

We note from the histograms that $\Delta\Delta G_{indir}$ makes a significant contribution to the total $\Delta\Delta G$ for all three inhibitors. The main contribution to this term for all 23 lysines comes from the Zn^{2+}

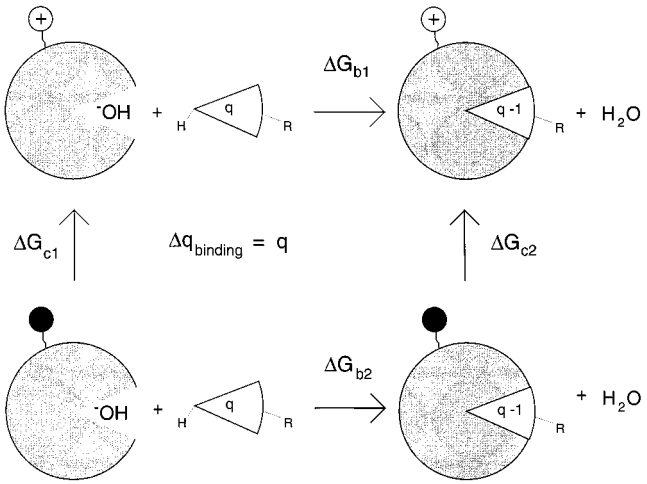


Figure 3. Thermodynamic cycle for the binding of a ligand to HCA II. HCA II is shown as a circle with a conical binding pocket. When it binds a sulfonamide ligand, hydroxide is displaced from the binding site. The lysine being acetylated is shown with a + charge in its native state and as neutral in its acetylated (uncharged) state. The difference in electrostatic binding free energy caused by acetylating the lysine can be determined by finding the energy of charging (deacetylating) the lysine in both the bound and unbound (hydroxide-bound) states.

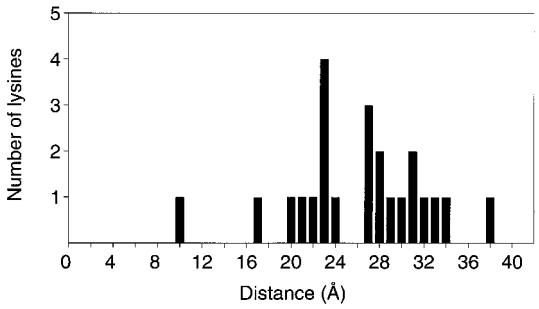
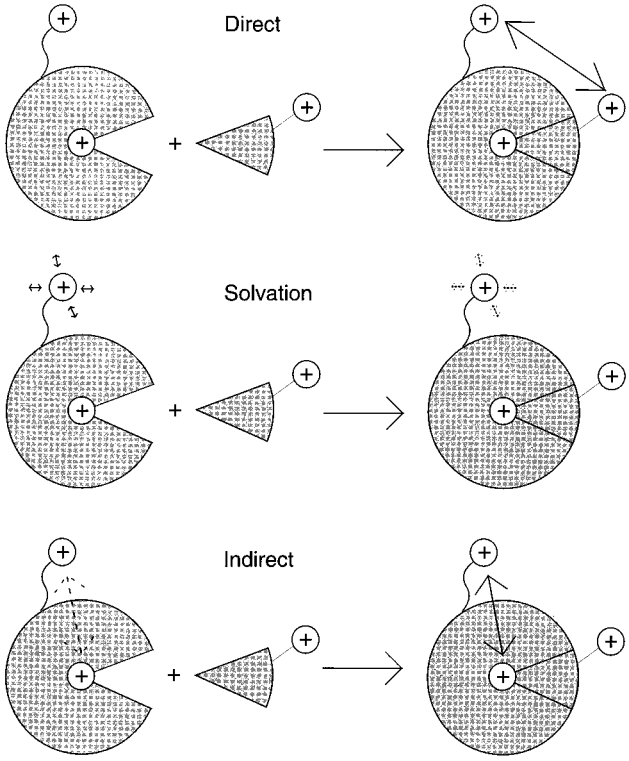


Figure 4. Histogram of distances from the Lys ϵ-amino groups in HCA II to the ammonium group of inhibitor 1 bound in the active site.

ion and the coordinated histidines. In the unbound state the Zn^{2+} ion is coordinated by a hydroxide ion, but in the bound state it is coordinated by a sulfonamide, whose larger low-dielectric region reduces solvent screening of Zn^{2+} interactions. It is an interesting question whether solvent in the roughly conical-shaped binding pocket is sufficiently oriented or restricted in motion that a dielectric constant substantially different from the bulk solvent value is required to simulate its effects. The values of $\Delta\Delta G_{indir}$ are similar among all three inhibitors, indicating that they are somewhat independent of the conformation of the R group on the benzene sulfonamide. (All three sulfonamide R groups have significantly different minimized conformations, although trial calculations suggest that the results are not particularly sensitive to the choice of conformation.)

In the case of (neutral) inhibitor 2, the average value of the total $\Delta\Delta G$ is essentially zero, as one might expect for a neutral group interacting with a charged group. However, a closer look at the histograms reveals that $\Delta\Delta G$ is made up of two non-zero terms that nearly cancel. $\Delta\Delta G_{indir}$ averages to +0.02 kcal/mol because the desolvation of the Zn^{2+} ion results in greater Zn^{2+}−Lys repulsion in the bound state than in the unbound state, but $\Delta\Delta G_{dir}$ averages to −0.02 kcal/mol. The Lys−sulfonamide interaction is more favorable than the Lys−hydroxide interaction, even though both hydroxide and sulfonamide have a total charge of −1, primarily because the sulfonamide's negative charge is less screened by solvent.

Scheme 2. An Illustration of the Three Contributions to $\Delta\Delta G$ of Binding



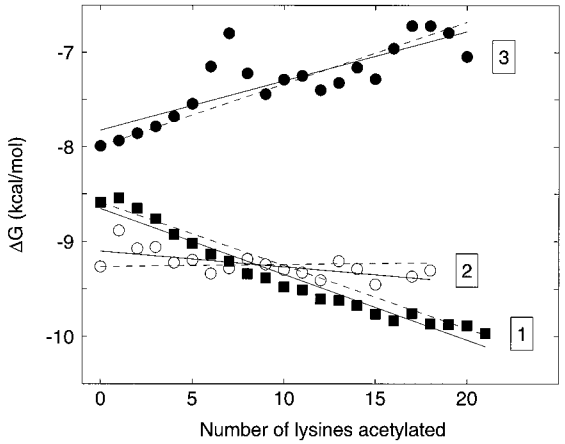


Figure 5. Plot of the free energy of binding (ΔG) determined by ACE *versus* the number of acetylated lysines (n) for each of the three inhibitors. The solid lines represent the least-squares fit to the data points, and the dashed lines represent the calculated values of the slopes as determined from the "null" model. The filled squares are the experimental data for inhibitor 1, the open circles are for inhibitor 2, and the filled circles are for inhibitor 3.

Discussion

Predicting a Charge Ladder Experiment from Calculated $\Delta\Delta G$ Values. A detailed, quantitative comparison of experiment and theory requires measuring or modeling the population of acetylation products in each rung of the charge ladder. Measurement of the relative reactivity of each lysine would be useful, for instance. In the absence of such information, the inverse question may be asked: is there a distribution of acetylations that produces the experimental population results from the individual binding calculations?

A full treatment of the possible patterns of acetylation in each rung allows a wide range of distributions. We made the

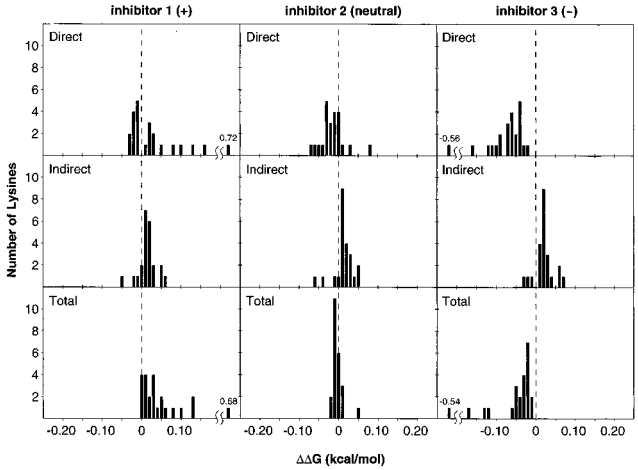


Figure 6. Histograms of $\Delta\Delta G_{dir}$, $\Delta\Delta G_{indir}$, and the total $\Delta\Delta G$ (including $\Delta\Delta G_{solv}$) as calculated for each of the three inhibitors. The values of $\Delta\Delta G_{dir}$ and the total $\Delta\Delta G$ calculated for Lys-170 in the presence of inhibitors 1 and 3 are much larger in magnitude than the other $\Delta\Delta G$ values. These four values are printed on the histograms.

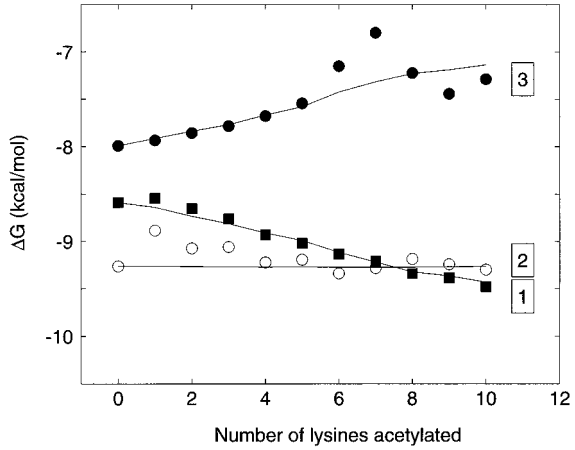


Figure 7. A simulated plot of ΔG *versus* n. The filled squares, open circles, and filled circles represent experimental data as in Figure 5. The lines represent a fit of the calculated $\Delta\Delta G$ values to the experimental data. The rates of acetylation of each lysine were chosen so that the fit to the experimental data was optimized.

following simplifying assumption to represent the vastly different possible patterns of acetylation. Let us define p_{ij} as the probability that lysine j is acetylated in going from the $(i-1)$th to the ith rung of the charge ladder. The set of p_{ij} values was optimized using a Monte Carlo algorithm so that the best possible agreement between ΔG_{expt} and ΔG_{calc} was obtained. The average deviation between ΔG_{expt} and ΔG_{calc} for only the first seven rungs of the charge ladders was used, since these ΔG values are expected to be the most accurate. Later rungs of the charge ladder have more uncertainty associated with them due to broadening of the peaks (particularly inhibitor 3). The results of the optimization are shown in Figure 7. We see that the overall trends in ΔG *versus* number of acetylations are reproduced, although the fluctuations in ΔG from one rung of the charge ladder to the next do not match exactly.

The results shown in Figure 7 can be reproduced using different sets of p_{ij}. To illustrate, we have plotted the fraction of each lysine that is acetylated in each rung of the charge ladder for 6 possible populations. In each part of Figure 8, the bottom line represents the fraction of each lysine that is acetylated in the second rung of the ladder. The second line in each part of

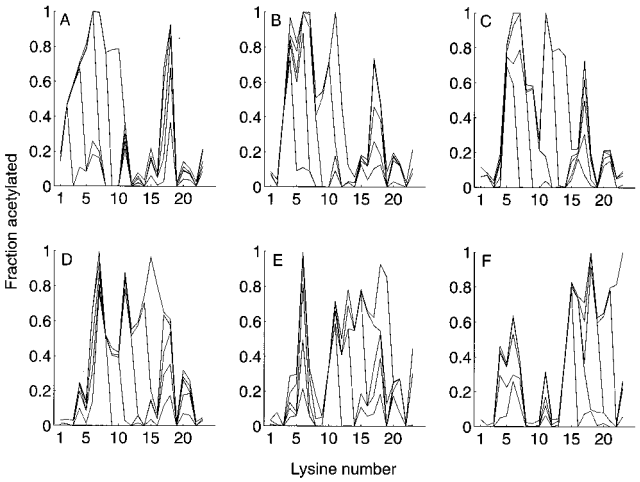


Figure 8. Six different patterns of acetylation that fit the data shown in Figure 7. In A−F, the bottom most line plots the fraction of each lysine that is acetylated in the second rung of the charge ladder. The second line from the bottom plots the fractions for the fourth rung, and so on. The patterns of acetylation are significantly different, yet they produce identical values of ΔG (see Figure 7).

Figure 8 represents the fraction of each lysine acetylated in the fourth rung, etc. (Every other rung is omitted from the figure for clarity.) We assume that this is the pattern of acetylation in the presence of all three inhibitors, since acetylation occurs prior to addition of any inhibitor. Although the patterns of acetylation are significantly different among the six parts of the figure, they all give identical predicted values of ΔG for each rung. There are thus multiple patterns of acetylation for which the calculated values of ΔG agree with the experimental values.

We may compare these results to those of the "null model", in which we assume that all lysines are acetylated at the same rate, and thus the observed $\Delta\Delta G$ would simply be an average of the calculated $\Delta\Delta G$ values for the lysines. These results are plotted in Figure 5. Here the slopes of the dashed lines are simply the average $\Delta\Delta G$ values for each of the three inhibitors, and the lines are drawn so that their intercepts are the ΔG values for rung 0 in each ladder. The difference in the experimental and calculated slopes is 0.00 kcal/mol for inhibitor 1, 0.01 kcal/mol for inhibitor 2, and 0.02 kcal/mol for inhibitor 3. These errors are on the order of experimental error, and therefore by this measure the agreement between theory and experiment is excellent, as seen in Figure 5.

Insensitivity to Outlying Values of $\Delta\Delta G_{calc}$. A striking feature of the distributions of $\Delta\Delta G_{calc}$ values for inhibitors 1 and 3 is that all lysines have $\Delta\Delta G$ values in the range 0.01−0.15 kcal/mol in magnitude except for one outlier (Lys-170), whose $\Delta\Delta G$ is much larger in magnitude. Would such a large $\Delta\Delta G$ be expected to cause a visible shoulder or peak splitting in the ACE results? We addressed this question by simulating a charge ladder experiment. Each protein derivative was assumed to make a contribution to the peak on the charge ladder as a Gaussian slightly narrower than the narrowest experimental line width. The derivatives were assumed to be present in the distribution predicted by the Monte Carlo method described above. The ligand concentration was chosen to maximize separation of the derivatives on the calculated charge ladder. The results are shown in Figure 9. Those derivatives acetylated at position 170 do not appear to separate from those not acetylated at 170, regardless of which distribution of protein derivatives is used. Although the charge ladder experiments provide an accurate measure of the average value for $\Delta\Delta G$ over

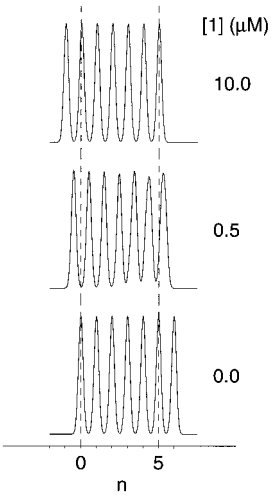


Figure 9. A simulated charge ladder experiment. The fraction of each possible protein derivative was chosen to be consistent with the ΔG *versus n* plot. The concentrations of inhibitor 1 are shown in the figure. 0.5 μM was chosen in order to maximize the separation between protein derivatives. Note that each rung appears as a single peak, as in the experiments, despite the spread in values of $\Delta\Delta G$ for each lysine. 10 μM is saturating, and the peaks are therefore shifted over one charge unit relative to their position with no ligand present.

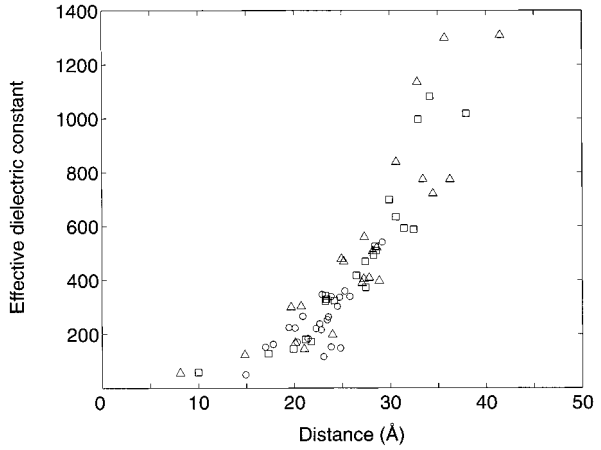


Figure 10. A plot of the effective dielectric constant for pairwise charge−charge interactions *versus* distance. The squares denote Lys−inhibitor 1 interactions, the triangles denote Lys−inhibitor 3 interactions, and the circles denote Lys−hydroxide interactions.

a large number of derivatives, they cannot rule out the existence of one or two outliers from the measured average $\Delta\Delta G$ in the large pool of derivatives unless higher resolution experiments are employed. Using our calculated values of $\Delta\Delta G$, splitting of peaks is expected when the experimental resolution is increased by a factor of roughly five.

Direct Interaction Strength as a Function of Distance. In Figure 10, the effective dielectric constant ($\epsilon_{\text{eff}} = q_1q_2/Er$ where ϵ_{eff} is the effective dielectric constant, q_1 and q_2 are the charges on the interacting groups, E is the screened Coulombic energy of interaction, and r is the distance between the groups) is plotted for each lysine interacting with the positive charge on inhibitor 1, the negative charge on inhibitor 3, and the negatively charged hydroxide. Clearly, the effective dielectric constant increases with distance. Other calculations have found that long-range interactions at distances of 10−15 Å usually have an effective dielectric constant of approximately 40.[3] These results agree with this value; however, the computed effective dielectric constant increases rapidly with distance at distances beyond 20 Å. Note that all of the interactions plotted in Figure 10 are between solvent-exposed charged groups. We might expect the effective dielectric constant to be substantially lower in cases where charges are buried.

Effects of Parameters on the Calculation. Decreasing the ionic strength from 25 mM to 15 mM increased the calculated total $\Delta\Delta G$ values almost uniformly by 13%. An increase in interaction energies is expected due to the decrease in screening by salt. When the ion exclusion radius was increased from 2.0 Å to 4.0 Å, the magnitudes of the total $\Delta\Delta G$ increased almost uniformly by 8%. Changing the interior dielectric constant from 4 to 3 resulted in an average change in $|\Delta\Delta G|$ of 3×10^{-3} kcal/mol; changing it to 6 resulted in an average change of 4×10^{-3}. The maximum change in the $|\Delta\Delta G|$ of any lysine caused by these changes in dielectric constant was less than 0.01 kcal/mol. Calculations were also performed with charges and radii from the CHARMM19 parameter set.[21] These calculations resulted in an average deviation from those made with the PARSE parameters of 7×10^{-3} kcal/mol. The results of the calculation are therefore not strongly dependent upon the choice of parameters.

Conclusion

The surface lysines of HCA II contribute to its binding of the family of benzene sulfonamide inhibitors studied here, primarily by long-range electrostatic interactions with partially buried active site groups. These interactions are of two types: direct interactions of lysine with ligand (sulfonamide in the bound state *versus* hydroxide in the unbound state) and indirect interactions of lysine with polar and charged protein groups in the active site that are differentially screened in the bound and unbound states. While most analyses of ligand binding focus on the former, both interactions are of similar magnitude here.

Experiments using affinity capillary electrophoresis provide aggregate information that is complementary to the detailed, though approximate, analysis with continuum electrostatic theory. A visual examination of the experimental electropherograms suggests a straightforward, uniform binding free energy contribution from each lysine of the protein because the rungs do not split at intermediate ligand concentration (partial saturation). The nearly constant increment in binding free energy per protein charge adds further support. One insight from the calculations is that such seemingly uniform behavior may be masking more complex underlying energetics that are homogenized through population averaging and instrument resolution. The use of higher-resolution capillary experiments will be helpful in examining the underlying distributions in more detail.

Experimental and Theoretical Details

Materials. Carbonic anhydrase II (human; pI 7.6; E.C. 4.2.1.1) was purchased from Sigma (St. Louis, MO). The protein, as purified from human erythrocytes, is acetylated at the N-terminal α-amino group. Inhibitor 1 was synthesized as described in ref 22; inhibitors 2 and 3 were synthesized as described in ref 23. Uncoated fused silica capillaries with an internal diameter of 50 μm were purchased from Polymicro Technologies (Phoenix, AZ).

Acetylation of Amino Groups of Proteins. Proteins were dissolved in water at a concentration of 0.1 mM, and 10 vol % of 0.1 N NaOH was added to each solution to bring the pH to 12. Five to twenty

(21) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(22) Gomez, F. A.; Avila, L. Z.; Chu, Y.-H.; Whitesides, G. M. *Anal. Chem.* **1994**, *66*, 1785.
(23) Avila, L. Z.; Chu, Y.-H.; Blossey, E. C.; Whitesides, G. M. *J. Med. Chem.* **1993**, *36*, 126.

equivalents of acetic anhydride (100 mM in dioxane) were added to the protein solution and the reactants were quickly mixed by vortexing. Reactions were usually complete within 1 min. Each rung in the final charge ladders was controlled to have similar intensity by mixing the products from several acetylations done with different equivalents of acetic anhydride. The sample was diluted in electrophoresis buffer (25 mM Tris, 192 mM Gly, pH 8.4) prior to analysis.

Capillary Electrophoresis (CE). CE experiments were conducted on a Beckman P/ACE 5500. Charge ladders of HCA II were analyzed at 25 °C on an uncoated capillary of fused silica (total length 47 cm, with a length to the detector of 40 cm) using 25 mM Tris-192 mM Gly buffer (pH 8.4) and an applied voltage of 15 kV.

Affinity Capillary Electrophoresis. The binding affinities of the proteins that make up the charge ladder of HCA II for the benzene sulfonamide inhibitors were measured using affinity capillary electrophoresis (ACE), as described elsewhere.[18] The binding constant for (neutral) inhibitor 2 was determined by competition with a charged ligand.

Structure of HCA II and Complex. Polar hydrogens were added to a crystal structure of HCA II complexed to a benzene sulfonamide inhibitor (PDB identifier 1cnw)[19] using the HBUILD facility in the program CHARMM.[21,24] The protein was held rigid in the unbound state in order to cancel grid energy terms from the solution of the linearized Poisson−Boltzmann equation. The root-mean-square deviation in C_α positions between the ligand-bound structure which was used and an unbound structure of HCA II[25] is 0.17 Å, confirming that there is very little structural change in HCA II upon binding a benzene sulfonamide. The sulfonamide structure was modified so that its R groups were those of ligands 1, 2, and 3. Those parts of the ligand structures that were not present in the X-ray crystal structure were minimized using the CHARMM22 all-atom parameter set.[26] The protein and sulfonamide group were fixed at their crystal structure coordinates during the minimization.

Partial Atomic Charges. Charges for HCA II (except the three histidine side chains coordinating Zn^{2+}, as described below) were taken from the PARSE parameter set.[27] The functional groups in the inhibitors that have PARSE parameters (e.g., amines, amides, and carboxylates) were assigned charges accordingly. However, other charges in the inhibitors (i.e., those of the benzene sulfonamide groups) were not given by the PARSE parameters and were therefore assigned by ab initio methods. The molecules were each divided into fragments separated by methylene groups and their charges assigned using the RESP method.[28] Geometry optimizations and electrostatic potential calculations were carried out with the program Gaussian 94[29] as described below.

First, the geometry of the benzene sulfonamide fragment of inhibitor 1 was determined at the 6-31G** level. The charged nitrogen and the carbon atoms bonded to it were omitted from the optimization, since their charges can be taken from the PARSE parameter set, and their orbitals would not be expected to significantly alter the charges in the rest of the molecule.

The geometry of zinc complexed with three imidazole rings (to model His 94, 96, and 119) and with a hydrogen-substituted sulfonamide was optimized at the 3-21G level with the Wachters−Hay all-electron basis set for zinc as implemented in the program Gaussian 94.[29] This smaller basis set was required to allow the calculation to finish in a reasonable time. Full optimization of this system resulted in a significant conformational change of the imidazole rings with respect to the crystal structure, so the complex was optimized while keeping the dihedral angles fixed to their X-ray crystal structure values. The difference in partial charges assigned with dihedrals free and fixed was less than 0.05 for all atoms except for zinc and the histidine nitrogens liganded directly to it. These atoms were more highly charged (each nitrogen by ca. −0.1 and zinc by +0.3) when the dihedrals were fixed.

The full benzene sulfonamide structure was appended to the zinc complex so that it had the proper orientation with respect to the zinc and the histidines. The electrostatic potential of the full complex was calculated at the 6-311G** level. Point charges were fit to the atomic centers using the RESP program, and the charges were constrained so that the total charge on the Zn^{2+} plus histidines was +2. The complex of hydroxide with the zinc and histidines was also assembled and optimized at the 3-21G level, again while keeping the dihedral angles along the Zn^{2+}−N coordination fixed. The charges on the Zn^{2+} and histidines were similar in both complexes (within 0.05 charge unit), and therefore their charges were averaged to give the charges used in the electrostatic calculations. The final charges used on the sulfonamides, the Zn^{2+}, and the histidines are shown in Scheme 1.

Continuum Electrostatic Calculations. Calculations of the electrostatic free energy were carried out using a locally modified version of the program DELPHI.[30−32] The linearized Poisson−Boltzmann equation was solved iteratively at 2.0 grids/Å and at an ionic strength of 0.025 M. Each calculation of the potential in the bound and unbound states was repeated with ten different translations of the grid with respect to the molecule (i.e., the molecular geometry was unchanged). The standard deviation of the $\Delta\Delta G$ values over these translations was never more than 5×10^{-4} kcal/mol, except in the case of Lys-170, where it was not more than 3×10^{-3} kcal/mol for any of the inhibitors. Therefore, the numerical error associated with the finite difference method is small compared to the values being computed. Trial calculations showed that Lys−Lys contributions to $\Delta\Delta G_{\text{indir}}$ could be neglected.

Acknowledgment. We thank Barry Honig for making the DELPHI computer program available and Martin Karplus for making the CHARMM computer program available. Figure 2 was created with QUANTA from Molecular Simulations, Inc. (San Diego, CA). This work was supported by the National Institutes of Health (GM51559 and GM30367 to G.M.W.; GM55758 and GM56552 to B.T.). J.A.C. is a National Science Foundation Predoctoral Fellow.

JA984195A

(24) Brünger, A. T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 148.

(25) Eriksson, A. E.; Jones, T. A.; Liljas, A. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 274.

(26) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

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

(28) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94 (Revision C.2)*; Gaussian, Inc.: Pittsburgh, PA, 1995.

(30) Gilson, M. K.; Honig, B. H. *Nature (London)* **1987**, *330*, 84.

(31) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327.

(32) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.