Calculations of Electrostatic Interactions and pK_as in the Active Site of *Escherichia coli* Thioredoxin[†,‡]

Valérie Dillet,[§] H. Jane Dyson, and Donald Bashford*

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received February 10, 1998; Revised Manuscript Received May 5, 1998

ABSTRACT: The active-site protonation state is crucial to the reductive mechanism of *Escherichia coli* thioredoxin, which involves a nucleophilic attack by the thiolate form of Cys32. We have calculated the titration properties of the active-site residues using a continuum electrostatic model, the X-ray structure of the oxidized protein, and ensembles of NMR structures of the oxidized and reduced protein. Protein dipoles, especially the SH dipole of Cys35, can provide sufficient stabilization of the Cys32 thiolate to account for its low experimental pK_a (∼7.4), but this effect is very sensitive to local conformational variations. The experimental finding that Cys32 titrates at a lower pH than Cys35 is explained by the latter's deeper burial from solvent exposure, and stronger interaction with the carboxylate of Asp26, and not by helix dipoles or positively charged side chains. The calculated very strong interaction between Cys32 and Cys35 in their thiolate forms implies that their titration must occur in two widely pH-separated steps and that the thiolate groups must move apart in the second step. The calculations are very consistent with the experimental Asp26 pK_a value of 7.5 for the oxidized X-ray structure. Both the oxidized and reduced NMR structures fall into two categories: "tight" structures in which the Asp26 and Lys57 side chains are in direct contact, and for which the calculations predict unreasonably low pK_as; and "loose" structures, which resemble the oxidized X-ray structure in that these side chains are farther apart, and for which the calculations are in very good agreement with experiment. We propose that the calculations over the NMR ensemble can be used as a test of the alternative structural models provided by NMR.

Thioredoxin is a protein disulfide reductase. Its catalytic activity strongly depends on the protonation state of the active-site region in the reduced form and involves unusual ranges of titration for at least two conserved ionizable residues: one of the two active-site cysteines and a nearby aspartic acid. In this study, we present electrostatic calculations of the pK_a shifts in the active site, compare them with experimental values, and analyze their causes. We examine how the results vary among a set of structures calculated from NMR[1] data, and we show that, in some cases, electrostatic models may help to distinguish those conformers which are most consistent with the acid−base properties of the protein.

The difference in stability between the reduced [dithiol, Trx-$(SH)_2$] form of thioredoxin and its oxidized (disulfide, Trx-S_2) form determines the powerful reductive properties of the thioredoxin system (thioredoxin and NADPH−thioredoxin reductase). This enzymatic system is involved in many biological processes (*1*, *2*), including the reduction of protein disulfides (*3*), ribonucleotide diphosphates (*4*), sulfate (*5*, *6*), and methionine sulfoxide (*5*, *7*). The active site, composed of two cysteines separated by a glycine and a proline (CGPC), is located at the N terminus of an α helix (*1*). The first step of the reaction mechanism proposed by Kallis and Holmgren (*8*) is a nucleophilic attack of one of the cysteines in its thiolate form on a disulfide bond of the protein substrate, forming a mixed disulfide intermediate. Since this mechanism implies the deprotonation of the attacking cysteine, the pK_a of this residue, and the pK_a of other titrating groups, which can influence the thiol−thiolate equilibrium, are very important for the function of the protein.

In *Escherichia coli* thioredoxin, the two cysteines of the active site are the only cysteine residues in the protein. Cys32 provides the reactive thiolate (*8*) and is more exposed to solvent, while Cys35 is buried (*9*, *10*). High-resolution X-ray crystallographic coordinates for the oxidized form (*9*) and NMR-determined solution structures of both the oxidized and the reduced forms are now available (*10*). The influence of two ionizable residues, Asp26 and Lys57, on the reactivity of the protein has been demonstrated by mutation experiments (*11*, *12*). In the oxidized form of the wild-type protein, Asp26 titrates with a pK_a of 7.5 (*13*, *14*). In the case of the reduced form, the analysis of the experimental data is complicated by the presence of three ionizable residues in the active-site region: Asp26, Cys32, and Cys35. NMR chemical shifts of the atoms of the cysteine side chains

[†] Supported by National Institutes of Health Grants GM45607 (D.B. and V.D.) and GM43238 (H.J.D.). V.D. is especially grateful to the LEDSS, the Department of Chemistry of Grenoble I University, and the French Ministry of National Education for supporting her stay at The Scripps Research Institute as a visiting scientist.

[‡] The PDB file names for the oxidized X-ray, oxidized NMR, and reduced NMR structures are 2trx, 1xoa, and 1xob, respectively.

[§] Permanent address: LEDSS, 301, rue de la Chimie, D. U. Saint Martin d'Hères-BP 53, 38000 Grenoble Cedex 9, France.

[1] Abbreviations: NMR, nuclear magnetic resonance; MEAD, macroscopic electrostatics with atomic detail; NOE, nuclear Overhauser effect; UV, ultraviolet.

S0006-2960(98)00333-X CCC: $15.00
Published on Web 06/25/1998

clearly show two apparent pK_as between pH 7 and 10, one around 7.5 and the other above 9 (*15*, *16*), but there have been conflicting interpretations. On the basis of Raman spectroscopy and the absence of NMR evidence of titration in the active-site region of the D26A mutant between pH 9 and 10, Wilson et al. assigned the pK_a higher than 9 in the wild-type Trx-$(SH)_2$ to Asp26 and concluded that both cysteines exhibit pK_as between 7 and 8 (*16*, *17*). A more direct measurement of the Asp26 titration in the wild type was made by Jeng et al. (*18*), who followed the Asp26 side chain ^{13}CO chemical shift as a function of pH and concluded that Asp26 titrates with a pK_a of 7.3−7.5 in the reduced form. Measuring the thiol ionization versus pH by UV absorbance experiments, they concluded that Cys32, like Asp26, titrates around 7.5, while Cys35 has a pK_a of 9.9 (*12*). More recently, Chivers et al. (*19*) argued that each apparent pK_a reflects a partial titration of both Cys32 and Asp26, whereas Cys35 is not deprotonated until the pH is above 10. For reasons described elsewhere (*12*, *18*), we believe that the interpretations assigning the lower pK_a values (7.0−8.0) to Asp26 in the reduced form are better supported by the experimental evidence, and we shall use these as the basis for comparisons with the calculations presented in this paper.

Continuum electrostatic models have proven to be useful for the calculation of pK_a shifts in proteins (*20*−*23*). For example, such a model was used to calculate mutation-induced pK_a shifts of the nucleophilic cysteine of DsbA, an enzyme which catalyzes the formation of protein disulfide bonds and is homologous to thioredoxin (*24*). The theory used is presented in the next section, which is followed by a section that describes the ensembles of oxidized and reduced structures used, and the computational details. The presentation and discussion of results will focus on the ionizable groups in and near the active site, the contributions of the different free energy terms of the theory to the results, and the effects of conformational variation across the ensemble.

THEORY

We assume that the difference between the free energy of protonating an ionizable group in a protein and that for the same group in a small model compound is caused by differences in the electrostatic work of altering the group's charges from their deprotonated to their protonated charge distribution in the protein, versus the model compound (Figure 1). We further assume that this work can be described adequately by an electrostatic model in which the protein (or model compound) is treated as a low-dielectric object with embedded charges and the solvent is treated as a continuum with the dielectric constant of bulk water (usually close to $\epsilon = 80$). The embedded charges and their locations are determined by the atomic partial charges and coordinates, and the boundary between the molecular and solvent dielectric regions is a complex surface determined by the atomic coordinates and radii. The effect of ionic strength is treated by a Debye−Hückel approach, leading to a linearized Poisson−Boltzmann equation. Here, we present an outline of the method and introduce some terms and notations to be used in the presentation of results. A more detailed description using a similar notation can be found in Bashford and Gerwert (*25*).

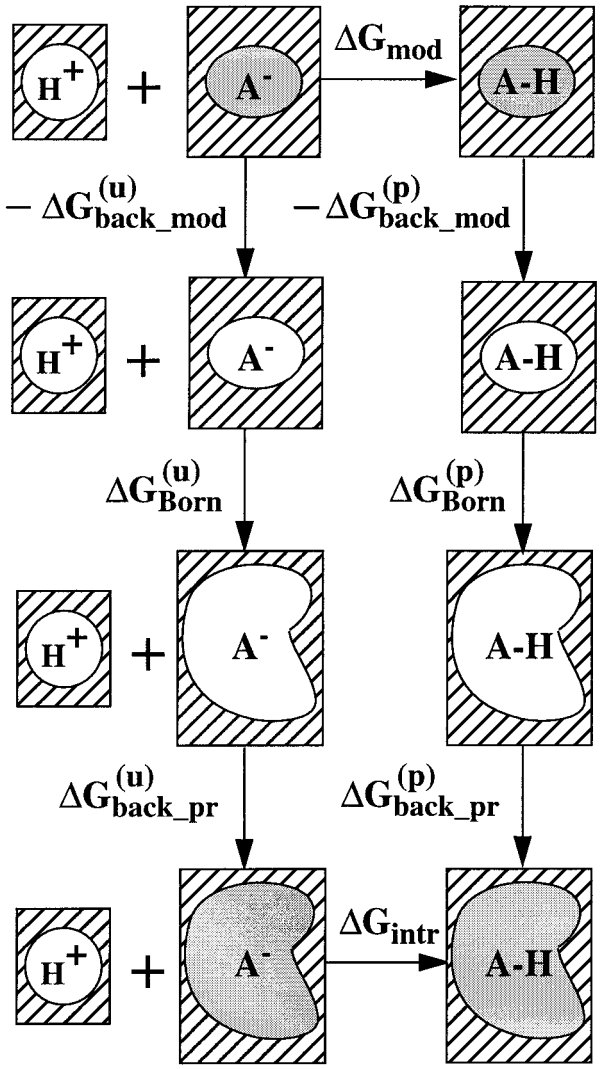

FIGURE 1: Thermodynamic cycle for the protonation of an ionizable group in a protein (last row), all the others residues assumed in their neutral form, with respect to the same process in a model compound (first row). The background and Born terms shown cause ΔG_{intr} to be different from the ΔG_{mod} value. Gray surfaces represent the presence of background charges in the protein or the model compound, whereas striped areas represent the solvent.

Our model implies that the titrating sites interact with each other in a pairwise additive way so that a complete theory of protein titration can be constructed in terms of the intrinsic pK_a of each site and the matrix of site−site interactions. The intrinsic pK_a (pK_{intr}) is defined as the pK_a a particular titrating group would have if all other titrating groups were held in their neutral form (*26*). In our electrostatic model, there are two kinds of effects contributing to the protonation free energy difference between the site in the neutralized protein and the model compound (Figure 1): the difference in the interaction of the titrating group's charges with the polarization that these charges themselves induce in the dielectric medium (Born solvation energy differences or $\Delta\Delta G_{Born}$) and the difference in the interaction of the titrating group's charges with the background of nontitrating charges and the charges of any other titrating groups in their neutral form ($\Delta\Delta G_{back}$). The expression relating pK_{intr} to the previously known pK_a of a corresponding model compound (pK_{mod}) is then

$$pK_{intr} = pK_{mod} - (2.303k_BT)^{-1}(\Delta\Delta G_{Born} + \Delta\Delta G_{back})$$

where k_B is the Boltzmann constant and T is the absolute temperature. The site−site interaction energy between two ionizable groups, i and j, is denoted by $W_{ij}z_iz_j$, where z_i and z_j are integers giving the formal charges of the groups.

To calculate the $\Delta\Delta G_{Born}$, $\Delta\Delta G_{back}$, and W_{ij} values pertaining to a particular site, i, the Poisson−Boltzmann equation for the electrostatic potential is solved for a set of problems in which the dielectric and electrolyte environments are determined by either the protein or model compound, and the charges generating the potential are those of either the protonated or deprotonated state of site i. The explicit formulas used are given by Bashford and Gerwert (*25*).

A protein molecule with N titratable sites has 2^N possible protonation microstates which we label by the variable ξ. Taking the reference state, ξ_0, to be the fully deprotonated state, the free energy of microstate ξ is

$$G(\xi) = \sum_{i=1} 2.303k_BTx_i(\xi)(pH - pK_{intr,i}) + {}^1/_2\sum_{i,j} W_{ij}z_i(\xi)z_j(\xi) - {}^1/_2\sum_{i,j}^{N} W_{ij}z_i(\xi_0)z_j(\xi_0)$$

where $x_i(\xi)$ is 1 or 0 according to whether site i is protonated or deprotonated in the protonation microstate ξ and $z_i(\xi)$ is the charge of site i in this microstate. The constant third term is added so that $G(\xi_0)$ is zero, but it has no influence on the final results. Finally, the fractional protonation of site i is obtained by a Boltzmann distribution over the ξ:

$$\langle x_i \rangle = \frac{\sum_{\xi}^{2^N} x_i(\xi)\exp[-\beta G(\xi)]}{\sum_{\xi}^{2^N}\exp[-\beta G(\xi)]} \quad (1)$$

where β equals $(k_BT)^{-1}$. In presenting our results, we will often express contributions to the free energy of protonation or deprotonation in terms of pK_a shifts. Specifically,

$$\Delta pK_{Born} = -(2.303k_BT)^{-1}\Delta\Delta G_{Born}$$

$$\Delta pK_{back} = -(2.303k_BT)^{-1}\Delta\Delta G_{back}$$

$$\text{influence of } j = -(2.303k_BT)^{-1}W_{ij}z_j$$

where the last expression is for a pK_a shift of site i due to the charged state of another site, j.

STRUCTURES AND METHODS

Calculations for the oxidized form of *E. coli* thioredoxin were performed using the X-ray crystallographic structure (*9*) and each of the 20 conformers from the NMR structural determination (*10*). For the reduced form, no X-ray structure is available, but we performed calculations on each of the 20 NMR-determined (*10*) conformers of the reduced molecule.

We used the CHARMM computer program (*27*) and the CHARMM22 empirical potential energy function (*28*) to prepare the PDB coordinate data for further calculations. Since our calculations use an all-hydrogen parameter set, and energetics of both protonated and deprotonated forms of all titrating groups were calculated, it was necessary to have coordinates for all hydrogen atoms, including those of the protonated forms of all titrating groups. The X-ray structure provided no hydrogen coordinates. The NMR structures, which were determined at pH 5.7, included all nontitrating hydrogen atoms, the hydrogen atoms of the carboxylic acid group of Asp26, and (for the reduced form) the thiol groups of Cys32 and Cys35 but did not include the carboxylic hydrogen atoms of the remaining aspartic and glutamic acid side chains. The HBUILD (*29*) procedure was used to add any missing hydrogen atom coordinates. All of the structures were then subjected to 200 steps of steepest-descent energy minimization without the electrostatic term, followed by a conjugate gradient minimization including electrostatic interactions (13.0 Å cutoff, dielectric constant of 4.0). All atoms were allowed to move during the whole minimization procedure. The all-atom root-mean-square deviation between original and minimized structures never exceeded 0.75 Å.

We have also carried out calculations on modeled structures of D26A and K57M mutants. These structures were prepared by modifying the wild-type NMR structures in the following way; for D26A, the carboxylic group of residue 26 was deleted, and for K57M, we replaced the C^{δ} atom of Lys57 with a sulfur atom and eliminated the H^{δ} atoms and the $N^{\zeta}H_3$ group. These structures were then minimized using the protocol described above.

The pK_{intr} and W_{ij} values pertaining to the titrating sites were computed using the multiflex program of the MEAD program suite (*25*, *30*), which uses the method of successive over-relaxation to solve a finite difference approximation of the Poisson−Boltzmann equation (*31*, *32*). Each evaluation of the electrostatic potential was carried out in two steps: the first using a finite difference lattice with a 1.0 Å spacing and the second with a 0.25 Å spacing. For the calculations in the protein environment and the model compound, the number of points along each edge of the cubic lattice was 81 and 61, respectively.

The dielectric boundary between the inner (protein or model compound) and the outer (solvent) media was defined by the contact and re-entrant surfaces (see ref *33*) using a solvent probe radius of 1.4 Å. Atomic radii and charges were taken from the PARSE parameter set of Sitkoff et al. (*34*). Dielectric constants of 4 and 80 were used for the solute (protein or model compound) and solvent, respectively. An ionic strength of 0.15 M was used together with an ion exclusion region extending 2.0 Å beyond the atomic radii. The model compounds used were the *N*-formyl-*N*-methylamide derivatives of the ionizable amino acids, and the pK_{mod} values corresponding to them were taken from Nozaki and Tanford (*35*): Asp, 4.0; Glu, 4.4; His, 7.0; Cys, 9.5; Tyr, 9.6; Lys, 10.4; Arg, 12.0; C terminus, 3.8; and N terminus, 7.5. As in previous calculations (*25*), the full sets of atomic partial charges were used for the protonated and deprotonated forms of the titrating groups. In other words, these were "distributed charge" calculations (*22*) rather than "single point charge" calculations.

The direct application of eq 1 for the protonation fraction, which requires summations over 2^N protonation microstates, is not practical for a protein having 32 (oxidized) or 34

Table 1: Calculated pK Values for the Active Site

	ΔpK_{Born}	ΔpK_{back}	pK_{intr}	pK_{half}
oxidized form				
Asp26 (X-ray)[a]	9.3	−1.5	11.8	7.4
Asp26 (loose)[b]	8.6	−1.3	11.3	7.7
Asp26 (tight)[c]	9.5	−2.6	10.9	2.2
reduced form				
Asp26 (loose)	8.6	−2.6	10.0	7.7
Asp26 (tight)	9.6	−3.8	9.8	0.9
Cys32 (full)[d]	3.7	−4.4	8.8	9.6
Cys32 (exc.)[e]	3.7	−4.8	8.4	9.2
Cys35 (full)	9.0	−7.8	10.7	>15

[a] Calculations on the X-ray structure (*9*). [b] Average results of calculations on the NMR structures which have longer Lys57−Asp26 distances (see the text). [c] Average results of calculations on the NMR structures which have shorter Lys57−Asp26 distances. [d] Average results of calculations on all 20 NMR structures. [e] Average results of calculations on all NMR structures except the four for which Cys35-H$^\gamma$ points away from Cys32-S$^\gamma$.

(reduced) titrating sites. Instead, a Monte Carlo method (*21*) was used. Protonation fractions of each site were calculated for pH values ranging from −4.0 to 15.0 in increments of 0.1 pH unit, and the pK_{half} value was defined as the pH at which the protonation fraction fell below 0.5.

RESULTS

The results of the calculations for Asp26 in oxidized thioredoxin (Trx-S_2) and for Asp26, Cys32, and Cys35 in reduced thioredoxin [Trx-$(SH)_2$] are summarized in Table 1. All of the other titratable residues were also included in the calculations, but results will only be presented for the active-site residues mentioned above and for Lys57 which has a strong influence on the active site.

Asp26 and Lys57 Titration

Reduced Form. The computed pK_{intr} and pK_{half} values of residues Asp26 and Lys57, in the 20 NMR structures of the reduced form of the wild-type protein, are displayed in Figure 2, along with distance and interaction data for this residue pair. The NMR structures fall into two distinct groups with respect to these results: a "tight" group (structures 7−20) in which the side chains of Asp26 and Lys57 are in van der Waals contact (Asp26-C$^\gamma$−Lys57-N$^\zeta$ distance less than 4.5 Å) and a "loose" group (structures 1−6) in which they are farther apart (the same interatomic distance greater than 5.5 Å).

In the case of Asp26, the pK_{intr} values fall in a relatively narrow range around an average of 9.8, but for the tight structures, the average pK_{half} of this residue is 0.9 while for the loose structures it is 7.7 (Table 1). The difference between the pK_{half} and pK_{intr} values, which in Figure 2A is the difference between the two bold curves, gives a measure of a residue's charge−charge interactions with other ionizable groups in their charged forms. As we will show presently, Lys57 exerts the principal charge−charge influence on Asp26, and vice versa.

In the case of Lys57, the pK_{intr} plot indicates two different average values: 8.1 for the loose structures and about 4.8 for the tight structures (Figure 2B). This difference is mostly due to a more negative ΔpK_{Born} term in the latter structures, which is caused by a more buried position of this lysine.

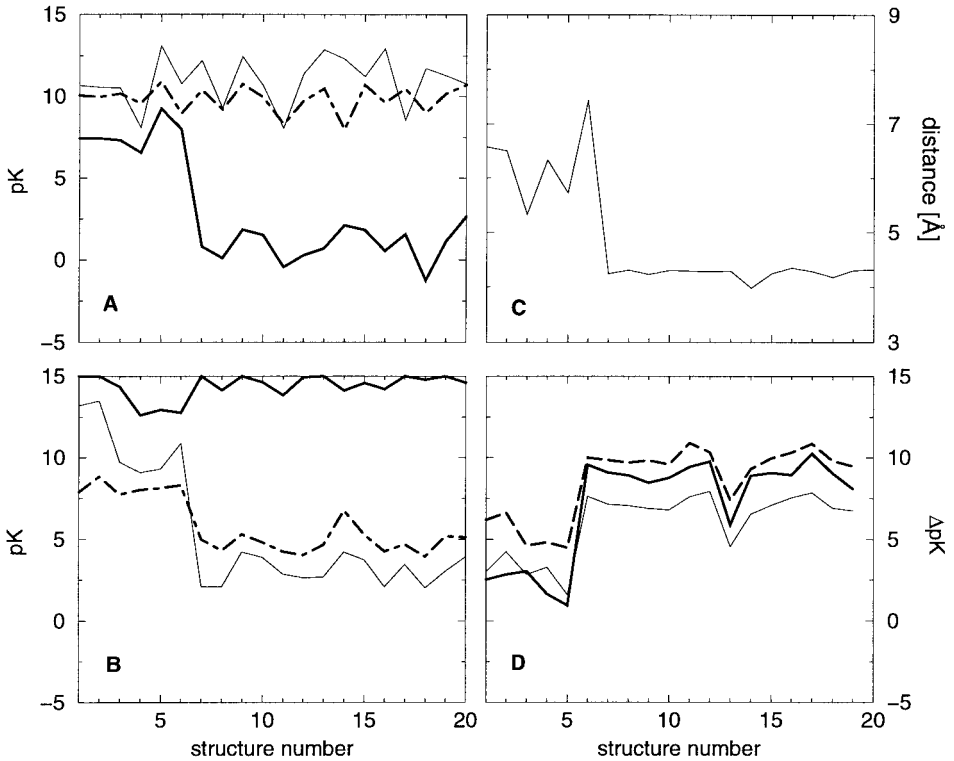

FIGURE 2: Calculated pK values or distances for the 20 NMR structures. The structures are numbered so that the loose structures are 1−6 and the tight structures are 7−20. (A) Calculated pK_{half} for Asp26 in the wild type (bold solid line) and in K57M (solid line) and pK_{intr} of Asp26 for the wild type (bold dot−dash line). (B) Calculated pK_{half} for Lys57 in the wild type (bold solid line) and in D26A (solid line) and pK_{intr} of Lys57 for the wild type (bold dot−dash line). (C) Distance (in angstroms) between Asp26-C$^\gamma$ and Lys57-N$^\zeta$. (D) Charge−charge interaction between Lys57 and Asp26 in pK units (solid line), $|pK_{half} - pK_{intr}|$ for Asp26 (bold solid line), and $|pK_{half} - pK_{intr}|$ for Lys57 (bold dashed line).

The pK_{half} values calculated for this residue display less variation, remaining above 12.5 for all structures. As was the case for Asp26, the difference between the two bold curves in Figure 2B shows a stronger interaction of this residue with other ionizable groups in the tight structures than in the loose structures.

The fact that the Asp26−Lys57 interaction is the main charge−charge interaction influencing both sites is demonstrated in Figure 2D, where the influence of this interaction can be seen to account for most of the pK_{half} − pK_{intr} difference for both residues. The correspondence of the variations in interaction to variations in the distance between these groups can be seen by comparing panels C and D of Figure 2.

In the K57M mutant, the p*K*-lowering effect of Lys57 on Asp26 is absent, and the pK_{half} values of Asp26 are unusually high (11.0 on average) and are similar to the pK_{intr} values (light solid line in Figure 2A). The mutation of Asp26 to Ala, in the D26A mutant, produces dramatic changes in the pK_{half} plot of Lys57 (light solid line in Figure 2B). For all the structures, the pK_{half} values are lower and lie closer to the pK_{intr} values compared to that of the wild-type protein, reflecting the loss of the charge−charge interaction with the aspartic acid. The tight structures give surprisingly low values for the pK_{half} of Lys57 (3.1 on average), whereas structures 1−6 lead to more realistic values (11.0 on average). This difference is again partly due to the more buried character of Lys57 in the tight structures compared to that in the loose structures, which causes a decrease in ΔpK_{Born} of about 3.3.

Oxidized Form. As shown in Table 1, the results obtained for the oxidized form of the wild-type protein using the 20 NMR structures are very similar to those reported above for the reduced form. For the four loose structures, the pK_{half} of Asp26 is close to 7.7, while for the remaining tight structures, the values are distributed around 2.2. As was the case for the dithiol form, this difference in the pK_a of Asp26 is correlated with a change (of about 1.6 Å) in the distance between Asp26-C$^\gamma$ and Lys57-N$^\zeta$, and with a change equivalent to 3.2 p*K* units in the site−site interaction between these two residues. In the crystallographic structure of the oxidized form, the distance between Asp26-C$^\gamma$ and Lys57-N$^\zeta$ is 6.1 Å, which is close to the average distance of 5.8 Å in the loose NMR structures. We obtain a pK_{half} value of 7.4 for Asp26 using the crystallographic structure.

Cys32

The average values of pK_{half} and pK_{intr} for Cys32 differ by only 0.8 (Table 1), and though both quantities fluctuate widely across the set of NMR structures, their difference never exceeds 1.3. This shows that the influence of site−site interactions on the pK_{half} value is relatively modest and consistent across the set of structures. The average influences of specific site−site interactions with the closest titratable residues (Lys36, Lys57, and Asp26) are reported in Table 2 together with the distance separating Cys32 from these groups. The strongest interaction is with the anionic residue, Asp26, while the interaction with the two lysines is weaker.

The value of the pK_{half} of Cys32 fluctuates between 6.5 and 12.7 in the set of structures. As shown in Figure 3A, ΔpK_{back}, the interaction with the nontitrating charges of the protein and other titratable groups in their neutral form, is the main term responsible for these fluctuations. (ΔpK_{Born} displays relatively little variation among the structures.) The contribution to the ΔpK_{back} of Cys32 due to the atomic partial charges of the Cys35 residue and the contribution of side chain charges only of this residue are also plotted in Figure 3A. These plots show that the fluctuations of the ΔpK_{back} values can be largely accounted for by the fluctuations of the Cys35 contributions, which in turn are largely fluctuations of the interaction with the SH dipole of Cys35. The distance between the sulfur atom of Cys32 and the thiol hydrogen atom of Cys35 for each of the 20 conformers is shown in Figure 3B. In all but four of the structures, this hydrogen atom is pointing toward the sulfur atom of Cys32. In the others, the hydrogen atom points away, and the electrostatic stabilization of the thiolate form of Cys32 due to the S−H dipole of Cys35 is decreased, resulting in a pK_a increase of about 1.5.

Table 2: Influence of Asp26, Lys36, and Lys57 on the Two Cysteines and Distances[a] between the Same Residues

	r_j		influence of *j*	
site *j*	Cys32	Cys35	Cys32	Cys35
Lys36	9.0	8.9	−0.3	−0.5
Lys57	10.7	7.2	−0.4	−1.8
Asp26	9.6	6.1	1.1	4.6

[a] Distances are in angstroms and refer to atom S$^\gamma$ for cysteines, C$^\gamma$ for the aspartic acid, and N$^\zeta$ for lysines.

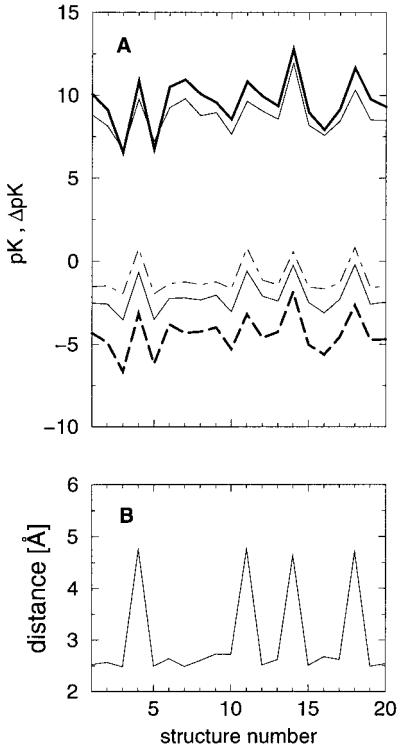

FIGURE 3: Calculated Cys32 p*K* values or distances for the 20 NMR structures numbered as in Figure 2. (A, top) pK_{half} (bold solid line) and pK_{intr} (solid line) and (A, bottom) ΔpK_{back} (bold dashed line), the contribution of Cys35 to ΔpK_{back} (solid line), and the contribution of the Cys35 side chain to ΔpK_{back} (dot−dash line). (B) Distance (in angstroms) between Cys32-S$^\gamma$ and Cys35-H$^\gamma$.

Table 3: Experimental pK_a Values

	reduced form			
	Asp26	Cys32	Cys35	oxidized Asp26
NMR	7.4[a]	7.4[b]	9.5[b]	7.5[a]
NMR[c]	7.5/9.2	7.5/9.2	>11	7.5
UV[d]	−	7.1	9.9	−

[a] Data from refs *13* and *18*. [b] Data from ref *15*. [c] Data from ref *19*. [d] Data from ref *12*.

Cys35

The pK_{intr} of Cys35 varies between 9.7 and 13.5 (average 10.7) for the different structures of the reduced form of the protein. The pK_{half} of this residue is always found to be higher than 15.0, indicating a strong coupling with negatively charged titrating groups. The strongest site−site interactions involving this cysteine are with Asp26, whose influence on Cys35 ranges from 3.3 to 6.2, and Cys32, whose influence ranges from 4.7 to 9.0. The average values of these interactions and the associated distances are shown in Table 2. Cys35 interacts much more strongly than Cys32 with Asp26, and thus, its titration is likely to be more sensitive to a change in the protonation state of this aspartic acid residue.

DISCUSSION

The calculations that we have presented are based not only on the assumptions of macroscopic electrostatics but also on the assumption that the protonation and deprotonation of side chains do not significantly alter the conformation of the protein. We expect that the latter assumption could lead to overestimates of some pK_a shifts since a conformational change in the real protein would provide a relaxation mechanism for countering large changes in electrostatic interactions. Therefore, we include considerations of possible conformational change in the interpretation of the calculations. In this work, the inclusion of a large number of conformers from NMR structure determination studies assists in the analysis of results by showing which aspects of the results persist across many conformers, and which are subject to large fluctuations. The calculations also suggest which of the NMR structures might be more representative of the protein in solution.

Titration of Active-Site Residues

The active site of *E. coli* thioredoxin contains three residues—Cys32, Cys35, and Asp26—for which experimental data on titration behavior are available (Table 3), but there is some controversy in the interpretation of these data, as outlined in the introductory section of this paper. A fourth ionizable residue, Lys57, is close enough to have strong electrostatic involvement. All four ionizable groups are buried to some extent; all are within 14 Å of each other, and some are much closer (see Figure 4). This leads to strong electrostatic effects, including desolvation effects (Born terms) and charge−dipole and charge−charge interactions. Thioredoxin therefore presents a challenging test of quantitative electrostatic calculations. Conversely, electrostatic calculations can give indications of the factors leading to the unusual pK_a values observed experimentally, even when small errors in large, opposing terms make it difficult to obtain very precise predictions of pK_a values.

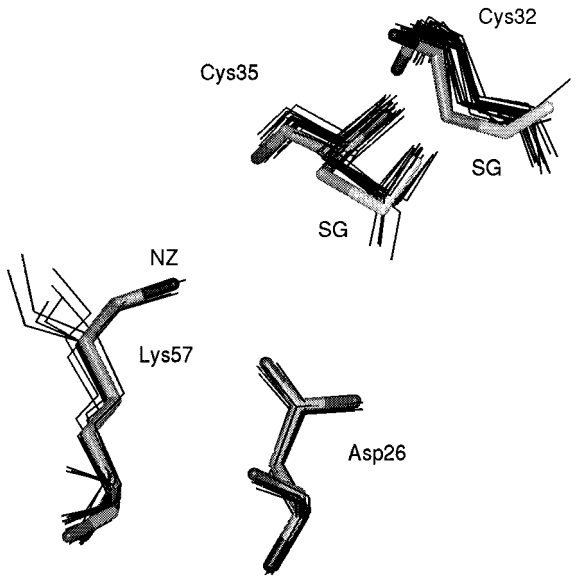

FIGURE 4: Active-site residues in Trx-$(SH)_2$. Only the heavy atoms and the thiol protons of the two cysteines are represented. All 20 NMR structures are displayed; the one displayed with a heaver stick is from the tight set. For this structure, the distances from Cys35-S^{γ} are as follows: Asp26-C^{γ}, 5.6 Å; Cys32-S^{γ}, 3.7 Å; and Lys57-N^{ξ}, 5.4 Å.

Comparison with Experimental Values. The experimental pK_as listed in Table 3 vary but are consistent on the following points: in the oxidized form, Asp26 titrates with a pK_a of 7.5; in the reduced form, both Asp26 and Cys32 titrate, at least partially, in the pH 7.0−8.0 range, while Cys35 titrates with a pK_a of 9.5 or higher. Our calculations for the loose NMR conformers and the X-ray crystallographic structure are consistent with the experiment for Asp26. For Cys32, our average calculated pK_{half} is considerably higher than 7.5, but for several NMR conformers, values near 7.5 are found. Chivers et al. (*19*) have proposed that Asp26 and Cys32 titrate in a coupled, two-step fashion such that the two groups compete for a single proton in the pH 8−9 range. Our individual-site titration curves for these groups (not shown) do not have the two-step structure that this coupled model implies. We do obtain a significant coupling (W_{ij} matrix element) between Asp26 and Cys32, but it ranges only from 1.1 to 1.7 kcal/mol, significantly smaller than the 2.4 kcal/mol coupling proposed by Chivers et al. These two groups are more than 9 Å apart in all structures, and since the tendency of calculations of the type we have performed is to overestimate interactions, we think it is unlikely that an electrostatic coupling as strong as 2.4 kcal/mol could occur between these groups. For Cys35, our calculations predict no deprotonation at pH values below 15, but some experimental data suggest that the pK_a is as low as 9.5 (*12*, *15*).

A more detailed understanding of the role of various active-site structural features, as well as some indication of the sources of deviation between experiment and theory, can be obtained by considering the energetic contributions leading to these calculated results.

Determinants of Cysteine pK Values. A major factor leading to low calculated pK_{half} values of Cys32 is the stabilization of the thiolate form of Cys32 by the thiol group of Cys35. This is demonstrated by the decomposition of

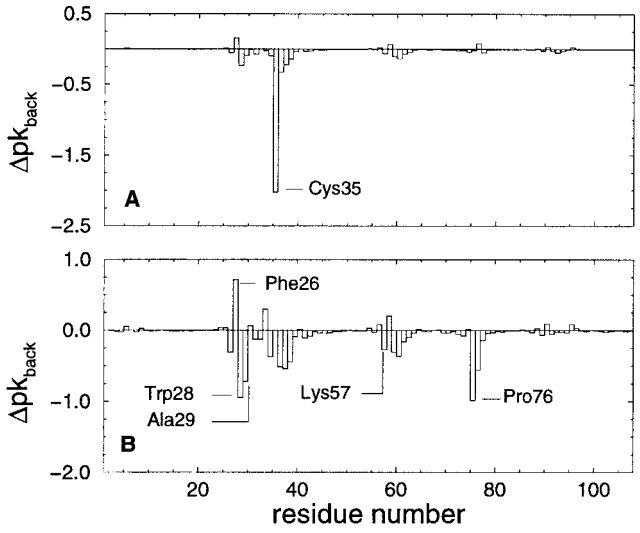

FIGURE 5: Electrostatic contribution to ΔpK_{back} of Cys32 (A) and Cys35 (B) of the microscopic dipoles of the other residues. The values reported in this graph represent the average over the 20 NMR structures of the wild-type *E. coli* thioredoxin.

the contributions of individual residues to ΔpK_{back} presented in Figure 5A. As illustrated in Figure 3A, many of the pK_{half} values calculated for Cys32 are too high, possibly indicating that this thiol−thiolate interaction is underestimated. Jeng et al. (*15*) have suggested that this interaction may have a strong "hydrogen bonding" character in which the proton is located between, and effectively shared between, the two sulfur atoms. This may lead to a stronger stabilizing influence than can be accounted for in the present calculations, which are strictly electrostatic, and rigid with respect to proton placement. In the remainder of this discussion, we will keep to the formalism where the proton is on either one cysteine or the other, since this is the formalism in which calculations were carried out, and it is usually the formalism for a discussion of thioredoxin's mechanism of reduction.

In all calculations on the reduced, wild-type form, the pK_{half} value of Cys35 is above 15, which is much higher than the value for Cys32. The short distance between the two cysteine sulfur atoms, 3.6−3.9 Å, leads to a strong electrostatic repulsion in the doubly deprotonated state, guaranteeing that a two-step titration will result, with one step having a significantly higher pK_a than the other. The rigidity of our model probably causes the effect to be overestimated, since it does not allow for an increase in the distance between the negative charges upon deprotonation of Cys35. Experimental interpretations assigning less drastic Cys35 pK_a values, such as 9.5 (*15*) or >11 (*19*), might therefore be reconciled with continuum electrostatic models if changes of a few angstroms in the sulfur−sulfur distance were allowed. The assignment of very similar pK_a values in the pH 7.0−8.0 range to both cysteines (*16*, *17*, *36*) is more difficult to accommodate, since the strong coupling between the two thiol groups necessitates two widely separated titration steps. A conformational change in the pH 7.0−8.0 range large enough to effectively decouple the two cysteines would need to be posited to support such an assignment. A conformational change of this size is hard to reconcile with the lack of any but extremely localized chemical shift changes in the NMR spectrum between pH 5.7 and ∼10.

The calculations present an opportunity to examine some earlier proposals regarding the stabilization of the Cys32 thiolate. Hol (*37*) suggested that the location of Cys32 at the amino-terminal end of the α_2 helix allows stabilization of the thiolate form by the helix dipole, and Kortemme and Creighton (*38*) concluded from model peptide studies that this effect could shift the pK_a of a cysteine residue by as much as 1.6. However, Jeng et al. (*10*) pointed out that the thiolate form of Cys35, which is also close to the amino terminus of the helix, should then be stabilized by a similar amount. Our finding that ΔpK_{back} is much more negative for Cys35 than for Cys32 suggests that the helix backbone interactions might be even greater for Cys35. As shown in Figure 5, Cys35 is involved in a greater number of favorable interactions with the dipoles of the protein than Cys32. The contribution of the α_2 helix backbone alone (residues 33−48) decreases the pK_{intr} value of Cys35 by 3.0 on average and that of Cys32 by 1.3 on average.[2] Helix dipole interactions, like the ΔpK_{back} terms in general, do not explain, indeed they contradict, the titration order of Cys32 and Cys35.

Repulsion between two thiolates explains why the titration of the cysteine pair should have two widely separated apparent pK_a values, but it does not explain why it is Cys32, and not Cys35, that titrates at the lower pK_a in the wild type. In our calculations, this is explained by the difference in ΔpK_{Born} terms (Cys35 is completely buried while Cys32 is partly exposed) and the difference in the interactions of Asp26 with Cys35 and Cys32 (Tables 1 and 2). The influence of ΔpK_{back} terms is in contrast to the ordering but is more than offset by the Born and site−site interaction terms. The involvement of Asp26 in ordering these pK_a values is consistent with the suggestion of Dyson et al. (*12*), based on their studies in which residue 26 is neutralized or mutated.

Kallis and Holmgren (*8*) suggested that positively charged residue(s) in the proximity of the active site, such as Lys36, could be partly responsible for the low pK_a value of Cys32. However, in the subsequently determined crystal structure (*9*), the side chain of Lys36 is too far from the sulfur atom of Cys32 (8.3 Å measured from the N^{ζ}) and no other positive residue is close. In our calculations, the pK_a shifts due to the charge−charge interactions between Cys32 and Lys36 or Lys57 are only 0.3 and 0.4, respectively, and no other positively charged group significantly interacts with Cys32.

Determinants of the Asp26 pK. The most extreme pK_a shift measured in the thioredoxin active site is that of the Asp26 side chain, which in the oxidized form has a pK_a value of 7.5 (*12*, *14*), 3.5 units higher than the model compound value. The calculations based on the X-ray structure of the oxidized form are very consistent with the experiment on this point (Table 1). In the calculations, this large upward shift in pK_{half} is caused mainly by the Born term (ΔpK_{Born} = 9.3) reflecting the unfavorable energetics of creating a charge in a buried environment. The Born term's influence is only partially offset by a favorable interaction between the negative carboxylate and the ammonium group of Lys57 and,

[2] In both cases, the interaction from the first two turns (residues 33−38) is responsible for more than 80% of the pK_a shift induced by the helix. Similar results were obtained from a theoretical study of helix dipole effects in barnase (*39*).

to a lesser extent, other dipolar protein groups. These general trends were also found in the calculations of Langsetmo et al. (*40*), who used the same X-ray structure, and a similar electrostatic model with different parameters, and who pointed out the consequences of such large pK_a shifts for the pH dependence of protein stability. However, the calculated pK_{half} of these workers was significantly greater than the experimental value, and they found better agreement with experiment if they made a rotation of the side chain dihedrals of Lys57 to bring its positive charge closer to the Asp26 carboxylate. In our calculations based on the loose NMR structures of the oxidized form, the results are very similar to those from the X-ray structure, not only in the final pK_{half} value but also in the relative contribution of different terms (Table 1). However, for the tight oxidized structures, we obtain pK_{half} values on average 5.2 units lower than the experimental values, primarily because of the dramatic increase in the strength of the Lys57−Asp26 interaction.

For the reduced form of the protein, only NMR structures are available, and the calculated results are very similar to those for the oxidized form. A large Born term, partially offset by the interaction with Lys57, leads to an average pK_{half} value of 7.7 for the loose structures, and an extremely low value for the tight structures. The implications of the difference between the calculations for the loose and tight structures will be discussed further below.

Conformational Variation

Using multiple NMR conformers for multiple calculations of an experimentally measurable quantity can provide benefits in two respects. The calculations can assist in the interpretation of structural data by suggesting which conformers are more plausible than the others (*41*), and the ensemble of NMR conformers can be used to assess the sensitivity of the computational methods to structural assumptions. We believe that the difference in the results for Asp26 in the loose versus tight structures provides an example of the former, and the variation in Cys32 results provides both kinds of benefits.

Loose versus Tight Conformers and the Asp26 pK_a. Most of the 20 conformers of oxidized thioredoxin in the NMR-derived ensemble (*10*) have an Asp26-C$^{\gamma}$−Lys57-N$^{\zeta}$ distance of less than 4.5 Å, and we have classified these structures as tight on the basis of this salt bridge-like geometry. In four of the conformers, this distance is greater than 5.5 Å, and we have classified these as loose. The situation for the reduced form is similar. Six of the 20 NMR conformers fall in the loose category, and the remainder are tight. Is one of these inter-side chain geometries more representative of thioredoxin in solution than the other? The X-ray structure of the oxidized form (*9*) more closely resembles the loose NMR conformers in the sense that the Asp26-C$^{\gamma}$−Lys57-N$^{\zeta}$ distance, in the X-ray structure, is 6.1 Å, and there is a water molecule between the two groups. In the NMR structure determination, no NOE cross-peaks were found between any side chain atoms of Asp26 and Lys57, and the 20 NOE bounds between the ϵ atoms of Lys57 and other nearby residues (Ile4, His6, Trp28, Ala39, and Asp43) are equally well satisfied in both the tight and loose structures. Since no water molecule between these residues was included in the refinement process, it is possible that the force field used in the refinement process tended to cause the lysine side chain to move into the empty space near Asp26 even though the aspartic acid was modeled in its neutral form and the force field had scaled-down formal charges (*10*). This might account for the majority of tight structures over loose ones, even though the NOE data did not discriminate between them.

We propose that our finding that the pK_{half} values calculated with the loose structures are consistent with experimental values, while those calculated from the tight structure are very far from experimental values, should be taken as evidence that the loose conformers are more representative of the true solution structure of thioredoxin in both oxidation states. For the oxidized state, there is the corroborating evidence of the X-ray structure. Although no X-ray structure is available for the reduced form, the general similarity of the high-resolution NMR structures of the oxidized and reduced forms suggests that structural details seen in the oxidized X-ray structure might also carry over into the reduced form since there is not NMR evidence to the contrary. In this connection, it would be interesting to re-refine the NMR structure with a water molecule present in the region of the Asp26 and Lys57 side chains. Such a refinement has been carried out for the human thioredoxin NMR structures (*42*).

Cys32 Fluctuations. The main source of the large variations in pK_{half} values calculated for Cys32 is the variation of the distance between the thiolate sulfur atom of Cys32 and the thiol proton of Cys35 over the set of NMR-determined conformers (see Figure 3 and the caption). This distance variation comes mostly from variation in the orientation of the Cys35 SH group, there being much less variation in the distance between the sulfur atoms of the two cysteine residues. No NMR signal can be detected for these two thiol protons (*10*), so the structure generation procedure placed them so that favorable hydrogen bonds would be generated. In the case of Cys35, a favorable interaction could be obtained by orienting the SH group toward either the backbone CO of Cys32 or the sulfur atom of Cys32, and the resulting ensemble of conformers contained instances of both orientations. In our calculations, the conformers with the Cys35 SH group oriented toward the Cys32 sulfur atom result in a lower Cys32 pK_{half} value, which is more consistent with the experiment.

It seems likely that, in the real protein, titration of the SH proton of either cysteine would induce reorientation of the other. In the pH range of 7.5−9.5 where Cys32 has lost its proton but Cys35 has not, the orientation of the Cys35 SH dipole toward the negatively charged sulfur would be more strongly favored than it would be at pH 5.7, where the NMR structure determination was carried out (*10*). This suggests that computational methods that allow for proton positions to change as the protonation state of the molecule changes would provide a better account of protonation energetics. You and Bashford (*43*) developed a method that couples individual site titrations with local conformational changes, but correlation of conformational change with more than one site's titration could not be included in the method. Development of methods to lift this restriction are underway (V. Spassov and D. Bashford, unpublished). Alexov and Gunner (*44*) recently developed a method that allows for these cross-

site correlations but allows only protons to move. The thioredoxin cysteines might be an interesting application of their technique.

Concluding Remarks

The work presented in this paper has shown that continuum electrostatic calculations can give a quantitative account of much of the unusual pH titration behavior in the active site of thioredoxin. The high pK_a of Asp26 results from a balance between the effect of burial, which destabilizes the charged form, and interaction with the positive charge of Lys57, which stabilizes it. The thiolate form of Cys32, which is essential for the nucleophilic attack in the thioredoxin mechanism (*8*), is stabilized by an interaction with the thiol hydrogen of Cys35. The preference for deprotonating Cys32 before Cys35 arises because of the greater solvent exposure of Cys32 and the greater influence of the negative charge of Asp26 on Cys35 than on Cys32.

Calculations carried out for each member of an ensemble of NMR-derived conformers have resulted in a proposal that one subset of these conformers, the loose subset in which Lys57 and Asp26 are not in immediate contact, is more likely to represent the solution structure of this region than other conformers. To explore this hypothesis, a re-refinement of the NMR structures using a model in which a water molecule can intervene between the two side chains could be carried out. The variations of results across the ensemble have also pointed to features that might have been missed in single-conformer calculations and have also highlighted the importance of developing improved methods of treating conformational change in such calculations.

COMPUTER PROGRAMS AND DATA

The MEAD computer programs, and the data files needed to reproduce a sampling of the results presented here, can be obtained over the Internet by anonymous FTP from the host ftp.scripps.edu in the directory pub/electrostatics.

ACKNOWLEDGMENT

We thank Dr. P. Hünenberger for his critical reading of the manuscript.

REFERENCES

1. Holmgren, A. (1985) *Annu. Rev. Biochem. 54*, 237−271.
2. Holmgren, A., and Björnstedt, M. (1995) *Methods Enzymol. 252*, 199−208.
3. Holmgren, A. (1979) *J. Biol. Chem. 254*, 9113−9119.
4. Laurent, T. C., Moore, E. C., and Reichard, P. (1964) *J. Biol. Chem. 239*, 3436−3444.
5. Porqué, P. G., Baldesten, A., and Reichard, P. (1970) *J. Biol. Chem. 245*, 2371−2374.
6. Tsang, M. L.-S. (1981) *J. Bacteriol. 146*, 1059−1066.
7. Brot, N., Weissbach, L., Werth, J., and Weissbach, H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2155−2158.
8. Kallis, G. B., and Holmgren, A. (1980) *J. Biol. Chem. 255*, 10261−10265.
9. Katti, S. K., and LeMaster, D. M. (1990) *J. Mol. Biol. 212*, 167−184.
10. Jeng, M. F., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E., and Dyson, H. J. (1994) *Structure 2*, 853−868.
11. Gleason, F. K. (1992) *Protein Sci. 1*, 609−616.
12. Dyson, H. J., Jeng, M. F., Tennant, L. L., Slaby, I., Lindell, M., Cui, D. S., Kuprin, S., and Holmgren, A. (1997) *Biochemistry 36*, 2622−2636.
13. Dyson, H. J., Tennant, L. L., and Holmgren, A. (1991) *Biochemistry 30*, 4262−4268.
14. Langsetmo, K., Fuchs, J. A., and Woodward, C. (1991) *Biochemistry 30*, 7603−7609.
15. Jeng, M. F., Holmgren, A., and Dyson, H. J. (1995) *Biochemistry 34*, 10101−10105.
16. Wilson, N. A., Barbar, E., Fuchs, J. A., and Woodward, C. (1995) *Biochemistry 34*, 8931−8939.
17. Li, H., Hanson, C., Fuchs, J. A., Woodward, C., and Thomas, G. J. (1993) *Biochemistry 32*, 5800−5808.
18. Jeng, M. F., and Dyson, H. J. (1996) *Biochemistry 35*, 1−6.
19. Chivers, P. T., Prehoda, K. E., Volkman, B. F., Kim, B. M., Markley, J. L., and Raines, R. T. (1997) *Biochemistry 36*, 14985−14991.
20. Bashford, D., and Karplus, M. (1990) *Biochemistry 29*, 10219−10225.
21. Beroza, P., Fredkin, D. R., Okamura, M. Y., and Feher, G. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5804−5808.
22. Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K., and Honig, B. (1993) *Proteins: Struct., Funct., Genet. 15*, 252−265.
23. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) *J. Mol. Biol. 238*, 415−436.
24. Warwicker, J., and Gane, P. J. (1996) *FEBS Lett. 385*, 105−108.
25. Bashford, D., and Gerwert, K. (1992) *J. Mol. Biol. 224*, 473−486.
26. Tanford, C., and Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333−5339.
27. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
28. MacKerell, A. D., Jr., Wiorkiewicz-Kuczera, J., and Karplus, M. (1995) *J. Am. Chem. Soc. 117*, 11946−11975.
29. Brunger, A. T., and Karplus, M. (1988) *Proteins: Struct., Funct., Genet. 4*, 148−156.
30. Bashford, D. (1997) in *Scientific Computing in Object-Oriented Parallel Environments* (Ishikawa, Y., Oldehoeft, R. R., Reynders, J. V. W., and Tholburn, M., Eds.) Vol. 1343 of Lecture Notes in Computer Science, pp 233−240, Springer, Berlin.
31. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) in *Numerical recipes in C. The art of scientific computing*, pp 673−680, Cambridge University Press, Cambridge, U.K.
32. Warwicker, J., and Watson, H. C. (1982) *J. Mol. Biol. 157*, 671−679.
33. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
34. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) *J. Phys. Chem. 98*, 1978−1988.
35. Nozaki, Y., and Tanford, C. (1967) *Methods Enzymol. 11*, 715−734.
36. Vohnik, S., Hanson, C., Tuma, R., Fuchs, J. A., Woodward, C., and Thomas, G. J., Jr. (1998) *Protein Sci. 7*, 193−200.
37. Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol. 45*, 149−195.
38. Kortemme, T., and Creighton, T. E. (1995) *J. Mol. Biol. 253*, 799−812.
39. Åquist, J., Luecke, H., Quiocho, F. A., and Warshel, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2026−2030.
40. Langsetmo, K., Fuchs, J. A., Woodward, C., and Sharp, K. A. (1991) *Biochemistry 30*, 7609−7614.
41. Khare, D., Alexander, P., Antosiewicz, J., Bryan, P., Gilson, M., and Orban, J. (1997) *Biochemistry 36*, 3580−3589.
42. Forman-Kay, J. D., Gronenborn, A. M., Wingfield, P. T., and Clore, G. M. (1991) *J. Mol. Biol. 220*, 209−216.
43. You, T. J., and Bashford, D. (1995) *Biophys. J. 69*, 1721−1733.
44. Alexov, E. G., and Gunner, M. R. (1997) *Biophys. J. 72*, 2075−2093.

BI980333X