Theoretical Study of the Unusual Protonation Properties of the Active Site Cysteines in Thioredoxin

Alexandra T. P. Carvalho, Pedro A. Fernandes, and Maria J. Ramos*

REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal

Received: June 17, 2005; In Final Form: January 27, 2006

The thiol/disulfide oxidoreductases of the thioredoxin family have, in the active site, two cysteines that can be in a reduced or an oxidized form. One of the cysteines in the reduced state is deprotonated, and it is called nucleophilic cysteine. The pK_a of this cysteine is different from that of a normal cysteine and varies widely among the different enzymes of this family. However, the factors responsible for the different degrees of stabilization of nucleophilic cysteine thiolate are not fully understood. Here, we have studied the well-known hypothesis of proton sharing between the active site thiols by performing a linear transit scan for the transfer of the proton between the active site cysteines. We used a two-layered (DFT/MM) ONIOM formalism, with the active site region treated at the B3LYP/6-31+G(d) level and the remains of the protein treated with the Amber Parm94 force field. The solvation free energy was accounted for with a continuum solvent model, by solving the Poisson−Boltzmann equation using the program Delphi. We have obtained excellent agreement with the experimental data available in the literature. Besides refuting the proton sharing hypothesis, our results include a value of 14.0 for the pK_a of the buried cysteine, a quantity that has not been possible to obtain experimentally but which has been proven to be higher than 11. Additionally, this study also provides detailed information on the very interesting and so far unknown fact that the contribution of the enzymatic structure (8.3 kcal/mol) prevails in relation to that of the solvent (0.60 kcal/mol) concerning the differential stabilization of the active site thiolates.

1. Introduction

The enzymes of the thioredoxin superfamily do not have a high degree of sequence similarity. However, they display a considerable degree of structural similarity. All of the enzymes have at least one domain with the thioredoxin fold.

The best characterized enzymes are thioredoxins, glutaredoxins, PDI, and DsbA. They all catalyze oxidation−reduction reactions. Thioredoxins and glutaredoxins function as protein disulfide reductases in the cytoplasm of eukaryotic cells; PDI belongs to the post-translational machinery of cells, being present in the endoplasmatic reticulum. PDI promotes the correct formation and isomerization of misformed disulfide bridges. Finally, DsbA catalyzes the oxidation of dithiols to disulfides and is present in the periplasm of prokaryotic cells.

These enzymes possess, in the active center, a characteristic CX_iX_jC motif, in which C represents a cysteine and X_iX_j two variable amino acids. These cysteines can be in an oxidized state, forming a disulfide bridge, or in a reduced state in the dithiol form, in which one of the cysteines (nucleophilic cysteine) is deprotonated and accessible to solvent and substrates and the other is protonated and buried.

Thioredoxin is a 12 kDa monomer with 188 amino acids. Near the active site, two conserved amino acids exist. One is Asp26, which is present in all thioredoxins, and the other, Lys57, is present only in prokaryotic thioredoxins.

The pK_a's of the active site cysteines have been measured many times with different methods, such as the pH dependence of the reaction with thiol reagents, Raman spectroscopy, and NMR analysis, and different interpretations have been proposed. The main difficulty in the interpretation of those measurements seems to arise from the proximity of three titrating groups (Cys32, Cys35, and Asp26).

Initially, it was proposed that the pK_a of the nucleophilic cysteine was smaller than the pK_a of a normal cysteine, which is 8.5 ± 0.5.[1] In the literature, a value around 7 is proposed for the pK_a of the nucleophilic cysteine and a value around 9 is proposed for the buried cysteine.[2,3] Subsequently, it was proposed that the two cysteines had a similar pK_a, lower than normal cysteines, with a value of 7.5.[4,5] More recently, from the results of an equilibrium study, Takahashi et al. proposed pK_a values for both cysteines of 9−10 and no linkage to other groups.[6]

In a most recent publication, Chivers et al.[7] brought some light to this issue by proposing that the titration value of thioredoxin can be explained on the basis of microscopic pK_a's for two interactive active site residues, Asp26 (7.5 and 9.2) and Cys32 (7.5 and 9.2). When the Asp26 is protonated, Cys32 has a pK_a of 7.5, but when the Asp26 is deprotonated, the pK_a of Cys32 increases to 9.2, due to the unfavorable interaction between the two negatively charged groups. The pK_a of Cys35 could not be determined, but it was shown to be higher than 11.

Actually, it is consensual that the different stability of the disulfide bond of the members of this family is not due to a difference in the bond itself but, instead, is a consequence of the different degree of stabilization of the thiolates of the nucleophilic cysteines. However, the factors that promote that stabilization are still poorly understood.

* Corresponding author. E-mail: mjramos@fc.up.pt.

10.1021/jp053275f CCC: $33.50
Published on Web 02/24/2006

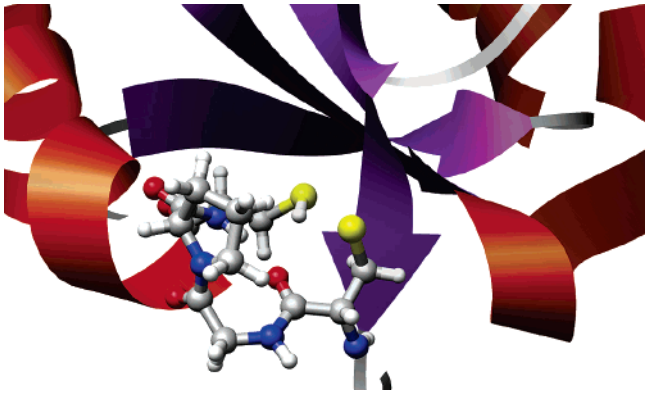

Figure 1. Active site of Trx taken from pdb 1XOB: the C*XX*C motif is shown in ball-and-stick representation, whereas the rest of the enzyme is shown in ribbon representation.

The factors pointed out in the literature are the variable amino acids in the CX_iX_jC motif,[8,9] the specific interactions with the residues in the vicinity of the active site cysteines,[10,11] the interaction with the macrodipole arising from the α2-helix,[12] a hydrogen bridge between the peptidic N of the buried cysteine and the S^{γ} of the nucleophilic cysteine, and proton sharing between the two sulfur atoms of active site cysteines.[3] This last hypothesis (proton sharing) was suggested on the basis of NMR observations, which indicated that the pK_a's of the thiol groups of the two active site cysteines of thioredoxin were shifted in opposite directions (about 1 pH unit) from that of a normal cysteine. The results were interpreted as an indication of strong interaction between the two groups manifested in proton sharing between the two sulfur atoms, by comparison with the behavior of dicarboxylic acids such as maleic acid.[13] The dependence of the chemical shifts and resonances of the H^{β} of the two cysteines with the pH also suggested a coupling and close interaction between the two cysteines.[3] Proton sharing between acidic groups has been inferred in the past in the active site of aspartic proteases,[13,14] myoglobin,[15] and bacteriorhodopsin.[16,17]

To test if proton sharing between the cysteines could indeed take place, we have performed a potential energy surface scan of the transfer of the proton between Cys32 and Cys35 in thioredoxin. From the obtained potential energy surface, the ΔpK_a between the two minima (protonated Cys32 and protonated Cys35) was calculated. We have analyzed also the influence of the solvent and the enzymatic structure in the stabilization of the thiolates.

2. Methods

The model system used was based on the NMR structure of reduced thioredoxin determined by Jeng et al.[18] (PDB entry 1X0B). From this structure, the proton in the nucleophilic S^{γ} atom of Cys32 was removed (Figure 1) and a shell of 2132 water molecules was added. Such a system is considerably large and has a small region which contributes mostly to the change in energy during proton transfer. Under such conditions, a layered theoretical treatment becomes the right choice. Accordingly, the ONIOM method was used to optimize the geometries and calculate the correspondent energies.[19] This method allows the division of a system in different layers, with each one being treated at a different theoretical level.

For the geometry optimizations, we have divided the system into two layers. The high level layer comprised the active site tetrapeptide Cys32-Gly33-Pro34-Cys35 (with the exception of the backbone of Cys32), corresponding to the CX_iX_jC motif. This layer was described at the density functional theory (DFT) level. The B3LYP functional was used. The 6-31G(d) basis set was employed, as implemented in Gaussian 03.[20] The low level layer was treated at the molecular mechanics level of theory, using the Amber Parm94 force field for proteins.[21] The water molecules were described by the TIP3P potential.

For the final energy calculations, a higher theoretical level was employed, the water molecules were deleted, and the system was divided into three layers. The high level layer comprised the residues Ala29, Cys32, Gly33, Pro34, Cys35, Ile75, and Pro76. This level was described at the DFT level with the B3LYP functional and the 6-31+G(d) basis set. The intermediate layer comprised the remains of the enzyme and was treated at the molecular mechanics level of theory using the Amber Parm94 force field.[21] The lower layer comprised the solvent, which was described through a dielectric continuum. Figure 2 illustrates the models.

Continuum solvent models allowed us to obtain the free energy of solvation, a quantity very difficult to obtain with discrete models. The solvation free energy was divided into two contributions, an electrostatic polar contribution (ΔG^{solv}_{polar}) and a nonpolar contribution ($\Delta G^{solv}_{nonpolar}$). The polar contribution was calculated by solving the Poisson−Boltzmann equation,[22,23] using the program Delphi.[24] The grid spacing was set to 0.4 Å, the molecule filled 80% of the grid box, the periodic boundary conditions were calculated with the option "full coulombic", and the convergence criterion was set to 0.001 kT/e. The dielectric constant of the protein was set to 3.0, and the dielectric constant of the solvent was set to 80.0. The ionic strength was set to zero. The default van der Waals (vdW) radii of Delphi and standard Amber Parm94 charges were used.

The nonpolar contribution, which includes the cavitation free energy and the vdW interactions between the solvent and the protein, was described by the following equation:

$$\Delta G^{solv}_{nonpolar} = \gamma A + b \qquad (1)$$

in which A is the solvent accessible surface area (calculated by the MSMS program)[25] and γ and b are two constants derived from experimental transfer energies of hydrocarbons between water/vacuum that have the values 0.005 42 kcal·mol^{-1}·Å^{-2} and 0.92 kcal/mol, respectively.[26] The probe radius was set to 1.4 Å.

The calculations were performed as follows: we began by optimizing the geometry of the reactants (i.e., thioredoxin with the Cys35 protonated and the Cys32 deprotonated) with standard Gaussian convergence criteria. Subsequently, we performed a relaxed potential energy surface scan, along the distance between the thiolate of Cys32 and the thiol proton of Cys35, approaching these two atoms with a step size of 0.07 Å, until the proton fully transferred to Cys32. Charges were derived for the transition state using the ESP fit and were linearly interpolated between the reagents and transition state and between the transition state and products.

The product has shown to be very hard to optimize freely, because (as expected experimentally by Chivers et al.[7]) the system was much more stable with the proton at Cys35, and the barrier for the reverse transfer was so low that it was overcome even in geometry optimizations. To avoid such a problem, a fine scan (step size of 0.01 Å) was performed around that minimum and the structure of the products was taken from the minimum of that scan. The difference in the energy of the products between the scan and freely optimized structure should be less than 0.3 kcal/mol.

A second smaller system was also constructed, comprising only two methanethiol molecules. This small model was only

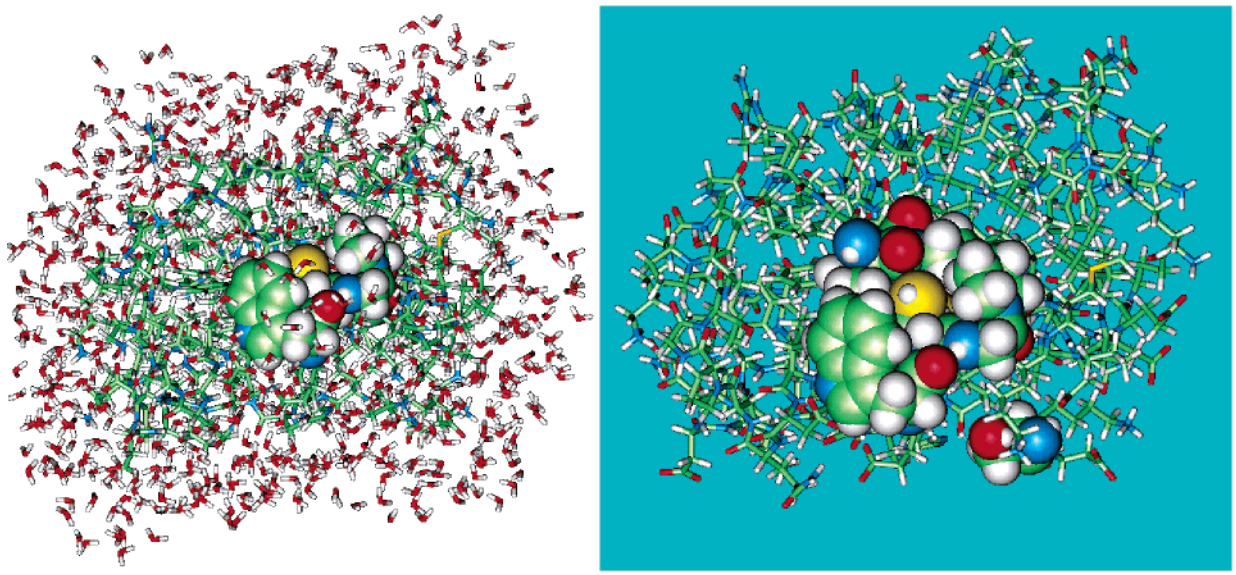

Figure 2. Models for the thioredoxin enzyme used in geometry optimizations (left) and energy calculations (right). The high level layers are represented in van der Waals representation, and the lower level in stick representation. The continuum is represented by the blue background. For geometry optimizations, the high level layer comprised the active site tetrapeptide Cys32-Gly33-Pro34-Cys35 (with the exception of the backbone of Cys32). In energy calculations, the high level layer comprised the residues Ala29, Cys32, Gly33, Pro34, Cys35, Ile75, and Pro76.

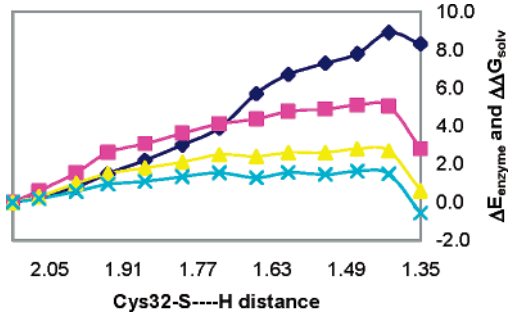

Figure 3. Relaxed potential energy surface scan of the proton transfer from Cys35 to Cys32 (the distances are in angstroms, and the energies are in kilocalories per mole): (◆) ΔE_{enzyme}; (■) $\Delta\Delta G_{solv}$ with $\epsilon_{enzyme} = 2$; (▲) $\Delta\Delta G_{solv}$ with $\epsilon_{enzyme} = 3$; (×) $\Delta\Delta G_{solv}$ with $\epsilon_{enzyme} = 4$.

used to estimate the contributions from the zero point energy, thermal correction and entropy to the activation free energy, and reaction free energy. These contributions were then extrapolated to the large system. This can be justified by noting that these contributions are very small and depend essentially on the vibrational term, which is very similar in the two systems, because it is essentially dependent on the type of bonds being broken and formed. This small system was treated at the DFT level of theory, using the B3LYP functional and the 6-31G(d) basis set. Full geometry optimizations of the reactants, transition state, and products of reaction were performed. Frequency calculations were carried out at all stationary points, with the number of imaginary frequencies confirming the nature of the stationary points.

3. Results

The potential energy surface displays three stationary states, one of them with an S−H distance to Cys35 of 1.42 Å, that clearly corresponds to a state in which Cys35 is protonated and Cys32 is deprotonated, and another with an S−H distance to Cys32 of 1.35 Å, now corresponding to a state in which Cys32 is protonated and Cys35 is deprotonated. In Figure 3, we have depicted the energy of the enzyme relative to the reactants (ΔE_{enzyme}) and the free energy of solvation, also relative to the reactants ($\Delta\Delta G_{solv}$); due to the small energetic barrier that separates the products from the previous state, it was necessary to perform a fine scan around the minimum obtained in the precedent scan. Detailed data relative to these scans can be found in Tables 1 and 2 in the Supporting Information.

The third stationary state has an energy above each one of the anterior states, corresponding to the transition state of the proton transfer. It corresponds to a late transition state, as expected from an endoenergetic reaction, with an S−H distance to Cys32 of 1.41 Å.

Several conclusions can be drawn from this potential energy surface; the first and more obvious is the impossibility of a scenario of proton sharing, because there is no minimum between the reactants and the products but instead a transition state. In addition, if we compare the contributions from ΔE_{enzyme} and from $\Delta\Delta G_{solv}$, we can observe that the major contribution to the preferential protonation of Cys35 comes from the protein environment and not the solvent (Figure 2), although Cys32 is partially exposed and Cys35 is buried. The only empirical parameter is the dielectric constant of the protein. Values between 2 and 4 are widely considered in the literature as the best choices. Although we have used the intermediate value ($\epsilon = 3$) in the calculations, which resulted in good agreement with the experimental data, we also checked the dependence of the results on this parameter and recalculated $\Delta\Delta G_{solv}$ for the other two values of the dielectric constant of the protein. We found that, even though the degree of exposure to the solvent was markedly different in the two cysteines, when we plotted $\Delta\Delta G_{solv}$ for the different values of the dielectric constant we obtained similar conclusions; hence, our results are not dependent on the dielectric constant chosen (Figure 3). It appears that the driving force for the differentiation of the degree of exposition of the two cysteines seems to be related to the catalytic mechanism, where it becomes assured that the substrate is limited to react with only one of the cysteines of the active center, and not with both. From the point of view of the differentiation of the basicity/nucleophilicity of the cysteines, the solvent does not play an important role.

The free energy for the stationary points, obtained by adding the zero point energy, thermal, and entropic corrections to ΔE_{enzyme} (extrapolated from the smaller model, see Table 3 in

the Supporting Information) and including also $\Delta\Delta G_{solv}$, is 12.5 kcal/mol for the transition state and 8.9 kcal/mol for the products. Considering a reaction free energy of 8.9 kcal/mol, we obtained a ΔpK_a value of 6.5, which is one more argument favoring the hypothesis of different pK_a's for the two cysteines. It is thus in disagreement with the experimental proposal of similar pK_a's for the two cysteines defended by both Li et al. and Wilson et al.[4,5] However, it is in strong agreement with the most recent proposal of Chivers et al.,[7] that defended the existence of different pK_a's for both enzymes, with the one for Cys35 being at least 3.5 units higher than the one for Cys32.

If we consider a value for the pK_a of the nucleophilic cysteine of about 7.5,[2−5,7,27−31] we obtain a pK_a of 14.0 for the buried cysteine, which is in agreement with the experimental results of Chivers et al., of a pK_a value higher than 11.[7]

The pK_a uncertainty in our calculations is around 1.4−3.2 pK_a units and arises from the normal errors associated with the energy calculation in the high level layer, the interaction energy between the two layers, and the solvation free energy. However, even considering the maximum value of error, the conclusions and their agreement with the experimental values are not affected.

4. Discussion

The most important conclusion taken from the relaxed potential energy surface scan is the impossibility of proton sharing between the two active site cysteines in thioredoxin.

If the proton was shared by the two active site cysteines, we should expect, when plotting the potential energy of the system with regard to the distance between the thiol proton and the S^{γ} of the nucleophilic cysteine, a curve with a minimum in the position where the proton was equally shared. Instead, we obtained a curve that reaches a maximum around 1.41 Å. This indicates that the proton is not shared but remains clearly attached to the buried cysteine in the most thermodynamically favorable position.

Another property that we have calculated is the difference in the pK_a between the two cysteines, which resulted in 6.5 pH units, leading to a pK_a of 14 for Cys35 if we assume a pK_a of 7.5 for Cys32,[2−5,7,27−31] in disagreement with older experimental proposals of similar pK_a's for the two cysteines[4,5] but in strong agreement with the most recent proposals that defended a pK_a for Cys35 at least 3.5 units higher than the one for Cys32.

In summary, our calculated potential energy profile allowed us to refute the proton sharing hypothesis. The obtained ΔpK_a value of 6.5 pK_a units predicts an increased value of 14 for the pK_a of Cys35. Finally, we proved that the difference in pK_a between the two cysteines is due to the different stabilization of the thiolates by the enzymatic structure, rather than by the solvent (although they have very different solvent accessibilities), which seems to indicate that the underlying reason for the different exposures of both residues is mechanistic, modulating the Cys−substrate interaction, rather than differentiating the fundamental difference in the acidic properties of both residues.

Acknowledgment. We are grateful for the financial support from Fundação para a Ciência e a Tecnologia (Portugal) for the fellowship SFRH/BD/17845/2004 and the National Foundation for Cancer Research (NFCR, U.S.A.).

Supporting Information Available: Tables showing relaxed potential surface scan point energies and free energies of solvation, energetic contributions for the total Gibbs free energy of activation and Gibbs free energy of reaction, and energies from the fine relaxed potential surface scan around the products. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Hoog, J. O.; Jornvall, H.; Holmgren, A.; Carliquist, M.; Persson, M. *Eur. J. Biochem.* **1983**, *136*, 223−232.

(2) Kallis, G. B.; Holmgren, A. *J. Biol. Chem.* **1980**, *255*, 261−265.

(3) Jeng, M. F.; Holmgren, A.; Dyson, H. *J. Biochem.* **1995**, *34*, 10101−10105.

(4) Li, H. M.; Hanson, C.; Fuchs, J. A.; Woodward, C.; Thomas, G. *J. Biochem.* **1993**, *32*, 5800−5808

(5) Wilson, N. A.; Barbar, E.; Fuchs, J. A.; Woodward, C. *Biochemistry* **1995**, *34*, 8931−8939.

(6) Takahashi, N.; Creighton, T. E. *Biochemistry* **1996**, *35*, 8342−8353.

(7) Chivers, P. T.; Prehoda, K. E.; Volkman, B. F.; Kim, B. M.; Markley, J. L.; Raines, R. T. *Biochemistry* **1997**, *36*, 14985−14991.

(8) Grauschopf, U.; Winther, J. R.; Korber, P.; Zander, T.; Dallinger, P.; Bardwell, J. C. A. *Cell* **1995**, *83*, 947−955.

(9) Gane, P. J.; Freedman, R. B.; Warwicker, J. *J. Mol. Biol.* **1995**, *249*, 376−387.

(10) Dyson, H. J.; Jeng, M. F.; Tennant, L. L.; Slaby, I.; Lindell, M.; Cui, D. S.; Kuprin, S.; Holmgren, A. *Biochemistry* **1997**, *36*, 2622−2636.

(11) Jacobi, A.; Huberwunderlich, M.; Hennecke, J.; Glockshuber, R. *J. Biol. Chem.* **1997**, *272*, 21692−21699.

(12) Kortemme, T.; Creighton, T. E. *J. Mol. Biol.* **1995**, *253*, 799−812.

(13) Pearl, L.; Blundell, T. *FEBS Lett.* **1984**, *174*, 96−101.

(14) Davies, D. R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189−215.

(15) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L. L.; Wright, P. E. *Biochem.* **1993**, *32*, 8045−8056.

(16) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473−486.

(17) Sampogna, R. V.; Honig, B. *Biophys. J.* **1994**, *66*, 1341−1352.

(18) Jeng, M. F.; Campbell, A. P.; Begley, T.; Holmgren, A.; Case, D. A.; Wright, P. E.; Dyson, H. J. *Structure* **1994**, 853.

(19) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, *461−462*, 1−21.

(20) *Gaussian 03*, revision B.04; Gaussian, Inc.: Pittsburgh, PA, 2003.

(21) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J. K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(22) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.

(23) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128−137.

(24) *DelPhi V.4*, release 1.0; Accelrys Inc.

(25) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.

(26) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(27) Dyson, H. J.; Tennant, L. L.; Holmgren, A. *Biochemistry* **1991**, *30*, 4262−4268.

(28) Forman-Kay, J. D.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1992**, *31*, 3442−3452.

(29) Li, H.; Hanson, C.; Woodward, C. K.; Thomas, G. J. *Biophys. J.* **1993**, *64*, A125.

(30) Qin, J.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1996**, *35*, 7−13.

(31) Dillet, V.; Dyson, H. J.; Bashford, D. *Biochemistry* **1998**, *37*, 10298−10306.