Direct Comparison of Experimental and Calculated Folding Free Energies for Hydrophobic Deletion Mutants of Chymotrypsin Inhibitor 2: Free Energy Perturbation Calculations Using Transition and Denatured States from Molecular Dynamics Simulations of Unfolding[†]

YongPing Pan and Valerie Daggett*

Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195-7610

Received September 19, 2000; Revised Manuscript Received January 9, 2001

ABSTRACT: Previous molecular dynamics (MD) simulations of thermal denaturation of chymotrypsin inhibitor 2 (CI2) have provided transition-state models in good agreement with experiment. Unfortunately, however, the comparisons have been necessarily indirect. The simulations have provided detailed structural information but not energetics, while from experiment, structure is inferred from a ratio of free energy changes upon mutation (Φ values). Here, direct comparison with experimental free energies is obtained by performing free energy perturbation calculations of hydrophobic deletion mutants of CI2 using transition- and denatured-state structures from various denaturation MD simulations. The agreement between the calculated and experimental $\Delta\Delta G$ and Φ values is quite good ($R = 0.8-0.9$). In addition, given the availability of realistic atomic models for the denatured protein, the common approach of using small peptides to represent the denatured state in stability calculations can now be evaluated. To this end, two different extended tripeptide models were used: one using the sequence from the protein with the residue to be mutated in the center and the other with this residue surrounded by Ala residues. The results for the two peptides agree neither with one another nor with the different full-length denatured-state models, which do provide results in good agreement with experiment. This finding is noteworthy because the denatured state of CI2 is very disrupted with little residual structure, such that the peptides might have been expected to serve as reasonable models. Overall the calculations presented here validate our previous MD-generated transition- and denatured-state models and therefore the simulated unfolding pathways and their relevance to refolding.

To elucidate the mechanism and pathway of protein folding/unfolding, it is necessary to characterize all steps along the way, including the transition state (*1*). Given that the study of transition states requires kinetic experiments, this state is guaranteed to be on the folding/unfolding pathway. Unfortunately, however, such transient high-energy states elude structural characterization by conventional techniques. As a result, information about transition states was of low resolution and phenomenological until the development of Φ-value analysis, via protein engineering, by Fersht and co-workers (*2*−*6*). With this method, single and multiple mutations are made throughout the protein of interest. Kinetic and thermodynamic experiments are then performed on both the wild-type and mutant proteins. Valuable information about the local structure around each residue is then inferred from the so-called phi value, Φ, which is a ratio of the destabilization free energies for mutation of the transition ($\Delta\Delta G_{TS-D}$) and native ($\Delta\Delta G_{N-D}$) states, respectively, relative to the unfolded, denatured state of the wild-type protein. In general, Φ takes a value between 0 and 1. A Φ value of 1 indicates that the transition state has nativelike extent of structure at the site of mutation. In contrast, a Φ value of 0 implies that the local structure in the transition state is similar to that of the unfolded state. A fractional value represents partial structure in the transition state. For hydrophobic deletion mutations, Φ reflects the ratio of contacts a particular residue makes with the rest of the protein in the transition state relative to the native, folded state (*3*, *7*), or in other words, the contacts lost upon mutation relative to the wild-type protein. Nonetheless, the complete three-dimensional structure of a transition state is still beyond the reach of experiment at present. Molecular dynamics (MD)[1] simulations, however, can provide detailed structural models of the transition state. By doing MD simulations, the entire trajectory along the folding/unfolding pathway can be generated. Putative transition states, as well as other intermediate states, can be obtained from such trajectories via a conformation clustering procedure (*7*, *8*). Unfortunately, however, one cannot determine free energies from these simulations. Instead, indirect comparisons with the experimental values for hydrophobic deletion mutations were made by evaluating the change in tertiary contacts of the mutated

[†] This work was supported by the National Institutes of Health (GM 50789).

* Corresponding author. Email: Daggett@u.washington.edu.

[1] Abbreviations: MD, molecular dynamics; FEP, free energy perturbation; CI2, chymotrypsin inhibitor 2; TS, transition state; RMSD, root-mean-square deviation.

10.1021/bi0022036 CCC: $20.00
Published on Web 02/08/2001

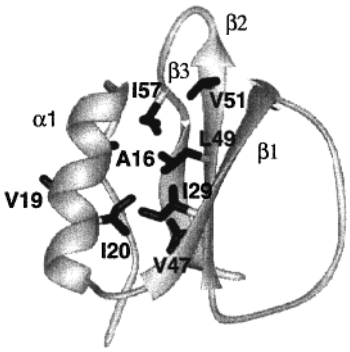

FIGURE 1: Ribbon representation of CI2 crystal structure (*18*). The side chains of residues mutated in this study are displayed and labeled.

residue of interest in both wild-type and mutant transition-state models to yield a Φ_{MD} value (*7*, *8*). Also, direct calculation of the local secondary and tertiary structure of each residue, relative to the native state, yields a structure index, *S*, for comparison with Φ values (*9*, *10*). While the agreement between the experimental Φ values and the calculated Φ_{MD} and *S* values has been quite favorable ($R = 0.8-0.95$), more direct comparisons of actual free energy changes upon mutation are desired.

In an effort to make direct comparisons with the experimental free energy changes, and resulting Φ values, we have performed free energy perturbation (FEP) calculations using transition- and denatured-state structures from the MD simulations. We have chosen to focus on the 64-residue protein chymotrypsin inhibitor 2 (CI2) (Figure 1). CI2 was the first protein shown to fold in a two-state manner by both kinetic and thermodynamic experiments (*11*). The study of the transition state of CI2 takes on special importance because its folding mechanism may be representative of other small proteins with a single domain or for individual domains of large proteins (*12*). Furthermore, many of the protein engineering procedures were worked out on CI2 (*11*−*15*), as it was the second protein subjected to such an analysis.

The first detailed high-resolution transition-state models for protein folding/unfolding by molecular dynamics were for CI2 (*7*, *8*), and we note that these studies were done as predictions, in parallel with the experiments. Furthermore, the procedures for identifying transition-state ensembles and their comparison with experiment (Φ_{MD} and *S* values described above) were developed on CI2. While the MD-generated models are in very good agreement with the experimental Φ values, we still desire an approach to directly calculate free energies for the conformational states sampled during MD. To this end, we have undertaken free energy perturbation (FEP) calculations of the hydrophobic deletion mutants investigated by Fersht and co-workers (*15*) (Figure 1). For a direct calculation of Φ values from the MD simulation, we have made the mutations in the native state, in transition-state models generated from three different MD simulations, and in five different denatured conformations after the protein fully unfolds. In addition, the availability of these detailed models of the denatured state allows us to test the validity of the more common approach of using tripeptides to model the unfolded state in calculations of protein stability. The mutations were made in each of the conformational states, and the corresponding free energy for the mutation was calculated, totaling approximately 100 free energy perturbation calculations. Through use of a thermodynamic cycle, as shown in Figure 2, these calculated free energies (in red) are then related to the experimental values (in blue). This is the first time such calculations have been performed on structures along an unfolding pathway.

MATERIALS AND METHODS

The FEP/MD method was used with a fixed window (*16*) as implemented within Amber 5.1 (*17*). The crystal structure (*18*) (Protein DataBank Accession No. 1YPC, 1.7 Å resolution, Figure 1) was used as the wild-type native state. The wild-type transition-state and denatured-state structures (Figure 2) were taken from previous simulations starting from both crystal and NMR structures (*7*, *8*). The TS structures (TS1, TS2, and TS3) are the 225, 345, and 180 ps snapshots, respectively, from the previous MD1, MD2, and a new MD simulation (Alonso et al., manuscript in preparation) at 498 K starting from the crystal structure. Five unfolded structures, D1a (3.0 ns) and D1b (2.9 ns) from the MD1 simulation, D2 (3 ns) from MD2, D3 (3 ns) from MD3, and D4 (3 ns) from MD4, were used to represent the denatured state, and they were obtained by continuing the previous MD simulations. All the systems were first solvated with a rectangular box of TIP3P waters (*19*), the edge of which was at least 10 Å from the nearest protein atoms. Then, the system was preequilibrated for 5 ps using the Sander module within Amber. Each FEP/MD calculation (the red arrows in Figure 2) was performed for 200 ps with 1000 windows, 100 steps each for equilibration and data collection, and a time step of 1 fs (a time step of 0.5 fs was used for some of the denatured-state calculations when 1 fs led to instabilities). All the simulations were performed at 298 K. The intermediate structures during the perturbation were generated using the single topology strategy. At the final stage of the mutation, dummy atom bonds were shrunk to 0.3 Å in length. The free energies calculated represent the interactions between the perturbed group and its environment; they do not contain internal energies for the perturbed group. The window growth method was used, based on an exponential formula, for the production FEP calculations described here, but the thermodynamic integration method gave comparable results in test calculations. The Cornell et al. (*20*) force field was used, and the nonbonded cutoff was 8 Å and was updated every 25 time steps. This short cutoff appears to be adequate for the type of mutation investigated here, which does not involve charged residues, and test calculations using a 10 Å cutoff produced similar results. SHAKE (*21*) was used to hamper motion of the covalent bonds. The extended tripeptide models, used for comparison with the full-length denatured protein, were acetylated (Ace) and amidated (Nme) at the N- and C-termini, respectively.

Approximately 100 calculations were performed on the I29V mutation in the native, TS, and denatured conformations to establish a protocol to ensure that coordinate changes during the free energy perturbation calculations were minimized and that the free energy changes along λ were smooth and minor between successive windows. In the test calculations, many protocols were tried with different combinations of window size and number of steps for each window. These test runs covered a time scale from 100 to 400 ps, with window numbers from 200 to 1500. Significant structural changes occurred with the TS models when the simulation

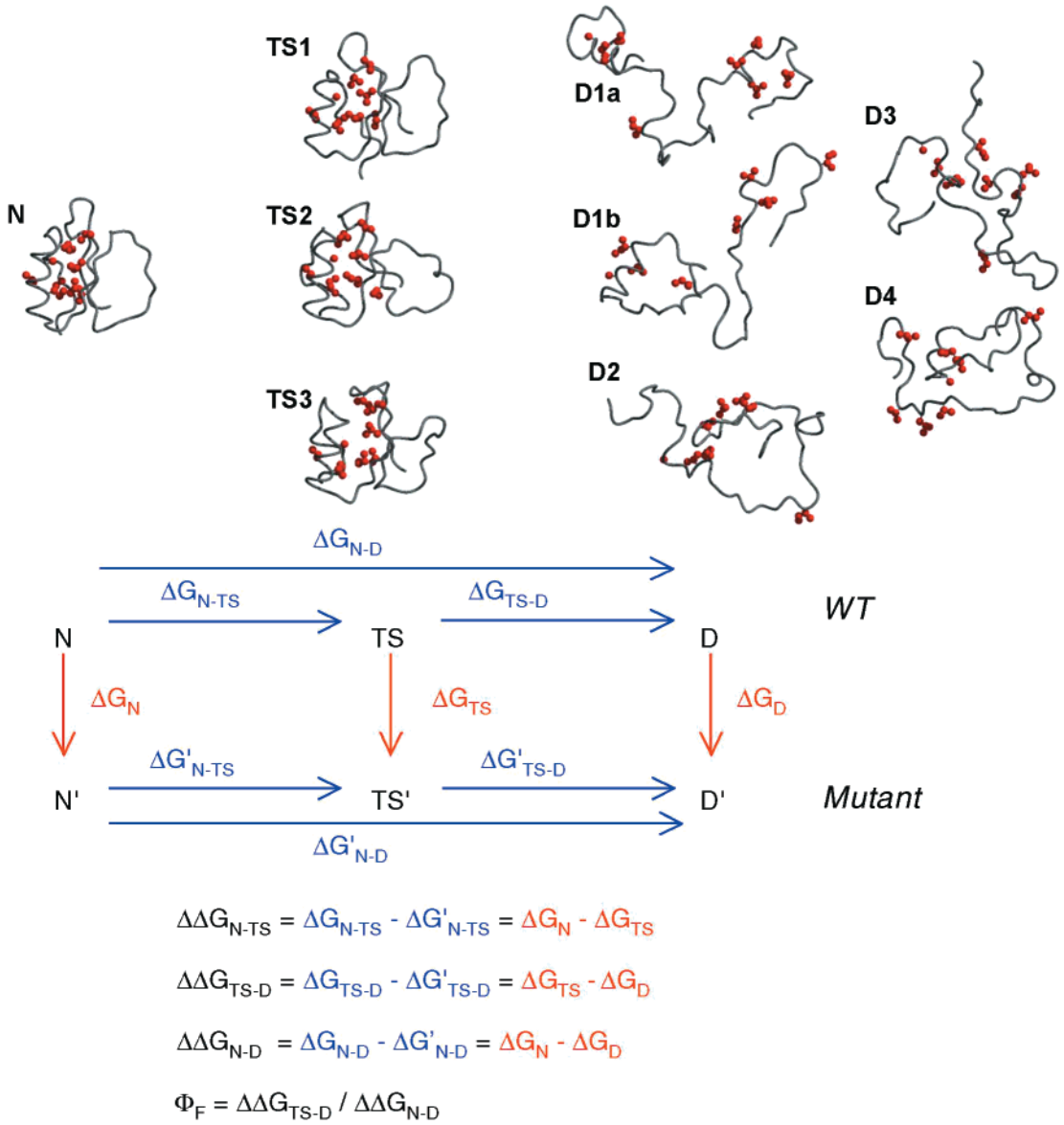

FIGURE 2: Thermodynamic cycles showing the relationship between the experimental and calculated free energies. N, TS, and D denote the native, transition, and denatured states of wild-type protein, and N′, TS′, and D′ are the corresponding mutants. The free energy changes in the horizontal direction (folding/unfolding) are available from experiment, and those in the vertical direction are calculated. Since free energy is a state function, the experimental and calculated free energies can be related as shown. The starting structures of CI2 used for the FEP calculations are displayed above the cycle with the residues to be mutated in red. The native structure (N) is the crystal structure. The transition-state (TS1−TS3) and denatured-state (D1−D4) structures are taken from simulations by Li and Daggett (*7, 8*) and Alonso et al. (manuscript in preparation).

time was longer than 250 ps, although the results for native and denatured states were still satisfactory. We note that the structural changes are expected since the transition state is transient and we also then quench it to 298 K, where it attempts to refold. An alternative is to run backbone-constrained simulations to avoid structural drift. However, with this approach different constraint forces gave quite different answers. Therefore, artifacts could be easily introduced by picking any one force constant. In the absence of constraints, simulations of the same length but with different window sizes yielded similar results. But we prefer to use a large number of windows, or small $\Delta\lambda$, since the mutation in each window is then smaller and the associated free energy change is smaller and smoother. Thus, we adopted the 200 steps per window for 1000 windows protocol to ensure a smooth structural change upon mutation and to minimize undesired structural changes of the TS.

At 100−200 ps, the free energy perturbation calculations are not very long although these times are standard for the field. Nevertheless, the short times could be problematic for the denatured-state simulations for which large conformational spaces are available and conformational sampling could be insufficient on this time scale. To test this idea, we ran 1 ns simulations on the I29V and A16G mutations for the QII and AAK peptide models for the denatured state. The peptides were chosen because deficiencies in sampling should be most evident in these small, mobile structures. The results of using different window sizes and different lengths of time for the calculations are illustrated in Figure 3. These results are also fairly independent of the number of windows used and number of steps at each window, with the exception of the 10-window protocol for AAK (Figure 3). The 200 ps and 1 ns simulations give similar results (Figure 3). Thus, we believe the protocol employed is appropriate in this

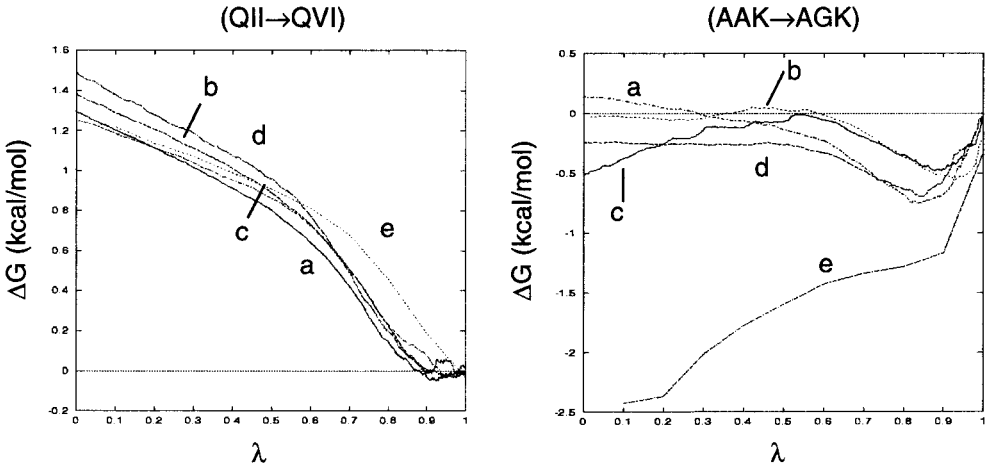

FIGURE 3: Free energy changes (ΔG) for two tripeptide models, QII and AAK, using different protocols. (a) 1000 windows with 100 steps each of equilibration and data collection for a total of 200 ps. (b) 200 windows with 2500 steps each of equilibration and data collection for a total of 1 ns. (c) 1000 windows with 250 steps each of equilibration and data collection for a total of 500 ps. (d) 1000 windows with 500 steps each of equilibration and data collection for a total of 1 ns. (e) 10 windows with 50 000 steps each of equilibration and data collection for a total of 1 ns.

Table 1: Calculated ΔG Values (kcal/mol) for TS and Denatured States[a]

mutant	D1a	D1b	D2	D3	D4	avg D	TS1	TS2	TS3	avg TS
A16G	1.27	0.95	1.16	0.41	1.69	1.10 ± 0.47	2.30	1.90	2.32	2.17 ± 0.24
V19A	0.98	1.35	1.34	1.17	0.65	1.09 ± 0.29	1.05	0.69	1.35	1.03 ± 0.33
I20V	1.44	1.26	1.39	1.51	1.50	1.42 ± 0.10	1.84	1.83	1.46	1.71 ± 0.22
I29V	1.30	1.32	0.43	1.36	0.79	1.04 ± 0.41	1.78	2.06	1.40	1.75 ± 0.33
V47A	0.32	1.04	1.22	0.68	0.45	0.74 ± 0.38	1.61	1.56	2.44	1.87 ± 0.49
L49A	0.26	0.70	0.82	0.61	0.78	0.63 ± 0.22	1.60	2.07	1.67	1.78 ± 0.25
V51A	2.20	2.36	1.75	1.79	1.49	1.92 ± 0.35	2.67	2.59	3.04	2.77 ± 0.24
I57A	1.96	0.98	2.30	1.48	2.28	1.80 ± 0.57	1.66	3.24	2.90	2.60 ± 0.83

[a] Denatured structures (D1−D4) and transition-state structures (TS1 and TS2) are taken from previous simulations (*7*, *8*). For example, D1a and TS2 are from MD1 and MD2, respectively. TS3 is from a more recent simulation by Alonso et al. (manuscript in preparation). All the denatured structures are the 3 ns snapshots from the appropriate trajectory, except D1b, which is the 2.9 ns snapshot of the MD1 simulation.

Table 2: Calculated and Experimental[a] $\Delta\Delta G$ (in kcal/mol) and Φ Values

				$\Delta\Delta G_{N-D}$		$\Delta\Delta G_{N-TS}$		$\Delta\Delta G_{TS-D}$		Φ	
mutant	ΔG_N	ΔG_{TS}	ΔG_D	calc	expt	calc	expt	calc	expt	calc	expt
A16G	2.28	2.17	1.10	1.18	1.09	0.11	−0.20	1.07	1.15	0.91	1.06
V19A	1.68	1.03	1.09	0.59	0.49	0.77	0.86	−0.06	−0.13	−0.10	−0.26
I20V	2.64	1.71	1.42	1.22	1.30	0.93	0.57	0.29	0.52	0.24	0.40
I29V	2.45	1.75	1.04	1.41	1.11	0.70	0.69	0.71	0.19	0.50	0.37
V47A	2.60	1.87	0.74	1.86	4.93	0.73	3.81	1.13	1.02	0.61	0.21
L49A	2.64	1.78	0.63	2.01	3.80	0.86	1.45	1.15	2.11	0.57	0.53
V51A	4.24	2.77	1.92	2.32	1.98	1.47	1.00	0.85	0.49	0.37	0.25
I57A	5.42	2.60	1.80	3.62	4.29	2.82	3.46	0.80	0.36	0.22	0.08
A16G/I57A[b]	7.19	4.73	1.09	6.10	7.33	2.46	3.93	3.64	1.91	0.60	0.31
I29V/I57A	9.20	6.76	3.79	5.41	−	2.44	−	2.97	−	0.55	−
I29A/I57V	8.02	4.56	2.57	5.45	4.08	3.46	2.33	1.99	1.18	0.37	0.39
correlation coefficient (R)[c]				0.82 (0.91)		0.61 (0.84)		0.73 (0.73)		0.85 (0.91)	

[a] Experimental data are from the folding/unfolding (*15*) experiments in water, which are believed to be the most reliable, and they are most closely suited to our conditions. The ΔG_{TS} and ΔG_D values are the averages from Table 1. [b] Only TS1, D1a, and D1b are calculated for the double mutations. [c] The correlation coefficients between the calculations and experiment were calculated for all mutants and excluding V47A, in parentheses.

particular application, even for the denatured state. The shorter calculations are in agreement with those obtained from longer simulations, and no serious sampling problems were found.

RESULTS AND DISCUSSION

The wild-type starting structures for the perturbation calculations are shown in Figure 2. The calculated free energy changes for each of the eight different single-site hydrophobic mutations (Figure 1) for the native (ΔG_N), transition (ΔG_{TS}), and denatured (ΔG_D) states are listed in Tables 1 and 2. Results for double mutants are also provided in Table 2. The calculated destabilization energies ($\Delta\Delta G_{N-D}$, $\Delta\Delta G_{N-TS}$, and $\Delta\Delta G_{TS-D}$) and the Φ values are summarized in Table 2. Both the $\Delta\Delta G$ and Φ values agree reasonably well, in general, with the experimental results, yielding correlation coefficients of 0.79 ($n = 30$, for comparison of $\Delta\Delta G_{N-D}$, $\Delta\Delta G_{N-TS}$, and $\Delta\Delta G_{TS-D}$ for 10 mutations, 1 of

Table 3: Solvent-Accessible Surface Area (SASA)[a] of Residues To Be Mutated and Structural Changes upon Mutation from Wild-Type Starting Structures

	N		TS1		D1a	
mutation	side chain SASA (Å^2)	Cα-RMSD (Å)	side chain SASA (Å^2)	Cα-RMSD (Å)	side chain SASA (Å^2)	Cα-RMSD (Å)
MD of WT		1.8		3.7		5.9
A16G	0	1.9	23	4.3	56	5.6
V19A	53	1.8	46	2.9	78	7.3
I20V	0	1.4	37	4.3	90	5.3
I29V	6	1.5	27	3.1	101	5.5
V47A	0	1.5	34	4.6	84	5.1
L49A	0	1.8	20	2.9	54	5.6
V51A	9	1.4	62	4.2	77	5.0
I57A	0	2.0	11	3.4	58	5.1
A16/I57A	0	1.5	35	3.6	114	4.5
I29A/I57V	6	1.4	38	3.5	159	4.1
I29V/I57A	6	2.0	38	3.8	159	6.3

[a] Solvent-accessible surface areas were calculated using NACCESS (*45*).

our double mutants is a prediction, so no experimental data are available) and 0.85 ($n = 10$, for comparison of Φ values for 10 mutants), respectively. The agreement becomes much better when V47A is excluded, yielding correlation coefficients of 0.88 ($n = 27$) and 0.91 ($n = 9$), respectively.

The agreement between our calculations and the experimental results for the A16G and V19A mutations is worth special attention. A16G is the only mutation to yield an experimental Φ value of 1. Our calculations are in good agreement (Table 2). The V19A mutation results in a negative Φ value, which suggests that this side chain makes more contacts in the transition state than in the native state. The Φ_{MD} value (*7*, *8*) based solely on the difference in contacts is in agreement with experiment, and our FEP calculation again successfully reproduces the experimental results (Table 2). Other mutations that delete one methylene group (I20V, I29V) are also in decent agreement with experiment (Table 2). On the other hand, the agreement worsens in some cases when the mutations involve more than two methylene groups. In the case of V47A and L49A, the calculated free energy changes are consistently much lower than the experiment results. The possible reasons for these discrepancies are discussed below.

Overall Structural Changes upon Mutation. One assumption of the protein engineering method as used to explore the TS structure of protein folding/unfolding is that the mutations do not significantly change the structures of the native, transition, or denatured states and that the protein will still follow the same pathway of folding/unfolding (*3*). That is, the mutations merely probe the structure of the wild-type protein. There is experimental evidence that the native state of CI2 can tolerate a variety of hydrophobic deletion mutations (*15*). Since the unfolding trajectories and the resulting transition-state ensembles are presumed to model the actual unfolding pathway and the real transition state, it is important to make sure that our FEP/MD calculations do not drastically alter the conformation of the protein. The Cα root-mean-square deviations (RMSDs) for the mutations of the native state compared with the wild-type starting structure are 1−2 Å (Table 3). These structural deviations are mainly due to the movement of the N- and C-termini and the active site loop (Figure 4), which are known to be mobile in the native state (*22*). Indeed, comparable structural changes (Cα-RMSD = 1.8 Å) were observed in a parallel 200 ps MD simulation using AMBER (Table 3). This finding implies that the conformational mobility of the protein or force-field-induced distortions are the major sources of the structural deviations. The Cα-RMSD values are larger for the transition and denatured states: 3−4 and 5−7 Å, respectively (Table 3). Given the unstable nature of the transition state, structural changes are expected to occur during the simulations. In addition, we note that the starting structures are taken from high-temperature simulations and they will also change in response to their new environment. Despite these changes, the mutant transition-state structures remain within the overall envelope of structures from independent wild-type MD simulations (Figure 4). The structural changes were even more drastic when the FEP/MD or just MD calculations were carried out longer (data not shown), which is why we limit the calculations to 200 ps.

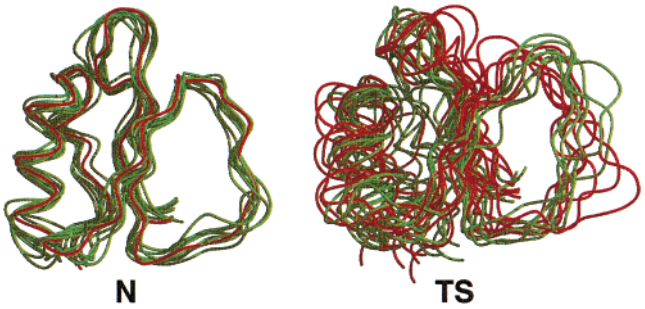

FIGURE 4: Superimposed structures of the wild-type (in red) and FEP calculated mutant (in green) proteins showing the structural similarity before and after the mutation for the native state (N) and the transition state (TS1). The wild-type transition-state structures were taken from molecular dynamics unfolding simulations by Li and Daggett (*7*, *8*).

Local Side Chain Rearrangement upon Mutation. The hydrophobic cores of the A16G, V19A, I20V, I29V, and V51A mutants are essentially undisturbed in the native state, and the calculated and experimental $\Delta\Delta G_{N-D}$ values are in good agreement, $R = 0.97$ (Table 2). Movement of the side chains occurs in the other mutants, for which the agreement between the experimental and calculated results is not as good. For the V47A mutant, I20 and I29 both move down toward the mutated residue (Figure 5). In the L49A mutation, A16 and I29 also move toward the mutation site, and I20 is pushed away (Figure 5). Similar side chain movements occur in the double mutants. The A16G/I57A mutations cause L49

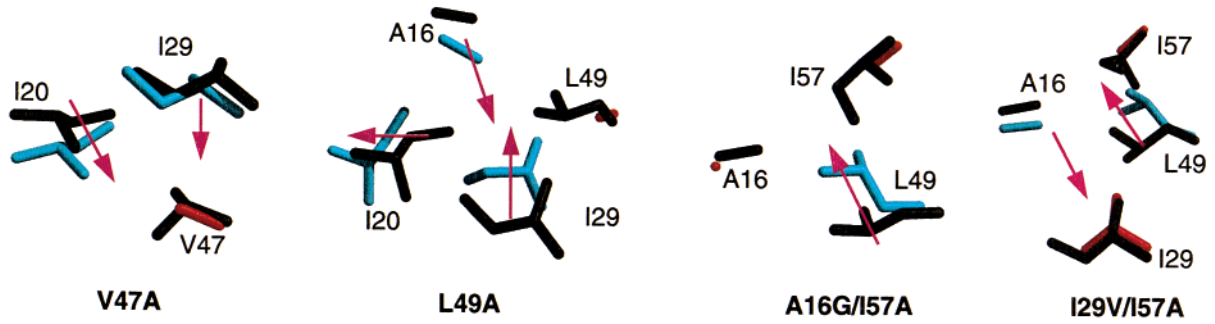

FIGURE 5: Superimposed wild-type and mutant residues showing the side chain rearrangement in the hydrophobic core of the native state upon mutation. Wild-type residues are in black, and mutants are colored cyan with the actual mutated residues in red. The arrows show the void created and the direction of the movement of the residues upon mutation.

to point toward the void introduced between I57 and A16 (Figure 5). In response, the rest of the hydrophobic core adjusts slightly. Similar movements are seen for A16 and L49 in the I29V/I57A double mutant. These structural rearrangements also occur in the transition state and appear to be responsible for the calculated underestimation of the destabilization upon mutation.

Nature of the Mutation and the Change in Free Energy. The relative free energies for a series of mutations obtained using the same protocol can provide insight into the relationship between the nature of the mutation and the magnitude of the free energy change. The native state yields free energy changes (ΔG_N) of 2.5, 4.2, 5.4, and 8.5 kcal/mol for deleting one (A16G, I20V, and I29V), two (V51A), three (I57A), and four (I29V/I57A, I29A/I57V) methyl/methylene groups, respectively (see Tables 1 and 2; V19A, V47A, and L49A are excluded for reasons discussed below). The larger the mutation (the number of CH_3- and/or CH_2-groups deleted), the larger the free energy change upon the native state. This effect is almost linear.

Given the structural changes observed at some positions, it is also of interest to compare the free energy changes for the same mutations at different sites. The ΔG_N for the V19A mutation is only 1.7 kcal/mol, while the values for V47A and V51A are 2.6 and 4.2 kcal/mol, respectively (Table 1). V19 is not as buried as V47 and V51, and therefore it makes fewer contacts with other residues. The solvent-accessible side chain surface areas for these residues are 53, 0, and 9 Å^2, respectively. In fact, the ΔG_N for removing two methyl groups for V19A is even less than for A16G, I20V, and I29V, which all involve deletion of only one methyl group.

To further investigate the relationship between the free energy change and the degree of hydrophobic contact, two double mutations were calculated for comparison: I29V/I57A and I29A/I57V. These two residues are both located in the hydrophobic core. Since the combined mutation volumes are the same, one would expect that they would give similar results. Interestingly, the difference in free energy between these two mutations is about 1.2 kcal/mol (Table 2), which is quite significant. The solvent-accessible surface areas for the I29 and I57 side chains in the native state are 6 and 0 Å^2, respectively. Given this information, it is reasonable to assume that I29 makes fewer contacts than I57 because part of its side chain is exposed to the solvent. Therefore, the I29V/I57A mutation should yield a greater free energy change than the I29A/I57V mutation, and, indeed, the calculations confirm this prediction, with values of 9.2 and 8.0 kcal/mol, respectively.

Comparison of Different Models of the Denatured State. While it is well-known that the effect of a mutation on stability reflects the difference between the native and denatured states, the effects are typically considered to be localized to the native state. As more evidence mounts that proteins can contain considerable residual structure in the denatured state (*23*−*28*), this approach must be questioned. In the case of CI2, however, the denatured state is nearly random coil (*29*−*31*). Nevertheless, even for this system ignoring the effects of mutation on the denatured state can be problematic. For example, ΔG_N for our series of mutants is somewhat correlated with $\Delta\Delta G_{N-D}$, $R = 0.68$, or 0.80 if V47A is excluded (Table 2). However, consideration of the effect of mutation on the denatured state by direct comparison of $\Delta\Delta G_{N-D}$ values improves the correlation; $R = 0.82$ and 0.91 for all the mutants and excluding V47A, respectively.

The denatured state of a protein is believed to be composed of many substates that are energetically similar to one another. Therefore, five unfolded structures from four independent MD trajectories were used to model the denatured state. These denatured structures are quite different in terms of Cα-RMSD values, ranging from 10 to 16 Å for any pair of them except between D1a and D1b, for which the Cα-RMSD is 6.6 Å. Interestingly, all models produced similar free energy results for each mutation (Table 1). This consistency reflects the tendency of these residues to make similar interactions with their environment irrespective of the overall structure of the protein.

The variance of the free energy changes for each individual mutation is again somewhat correlated to the environment of the mutation site. Taking D1a as an example, comparison of the local environment of V19, V47, and V51 reveals that V51 makes some favorable hydrophobic interactions with neighboring residues while the other two are much more exposed to the solvent in the denatured structures (Figure 2). This effect accounts for the low free energy changes for V19A and V47A (0.98 and 0.32 kcal/mol, respectively) and the significantly higher free energy change for the V51A mutation (2.20 kcal/mol) (Table 1). The free energy change for the I29V/I57A double mutation in D1a is larger than that of the I29A/I57V mutation (Table 2), where the side chain solvent-accessible surface areas for I29 and I57 are 101 and 58 Å^2, respectively (Table 3). Thus, the relative extent of solvent exposure and hydrophobic interactions for a residue with its surrounding environment account for the differences in the free energies for these mutations.

These are the first free energy perturbation calculations using all-atom models for the denatured state. In previous

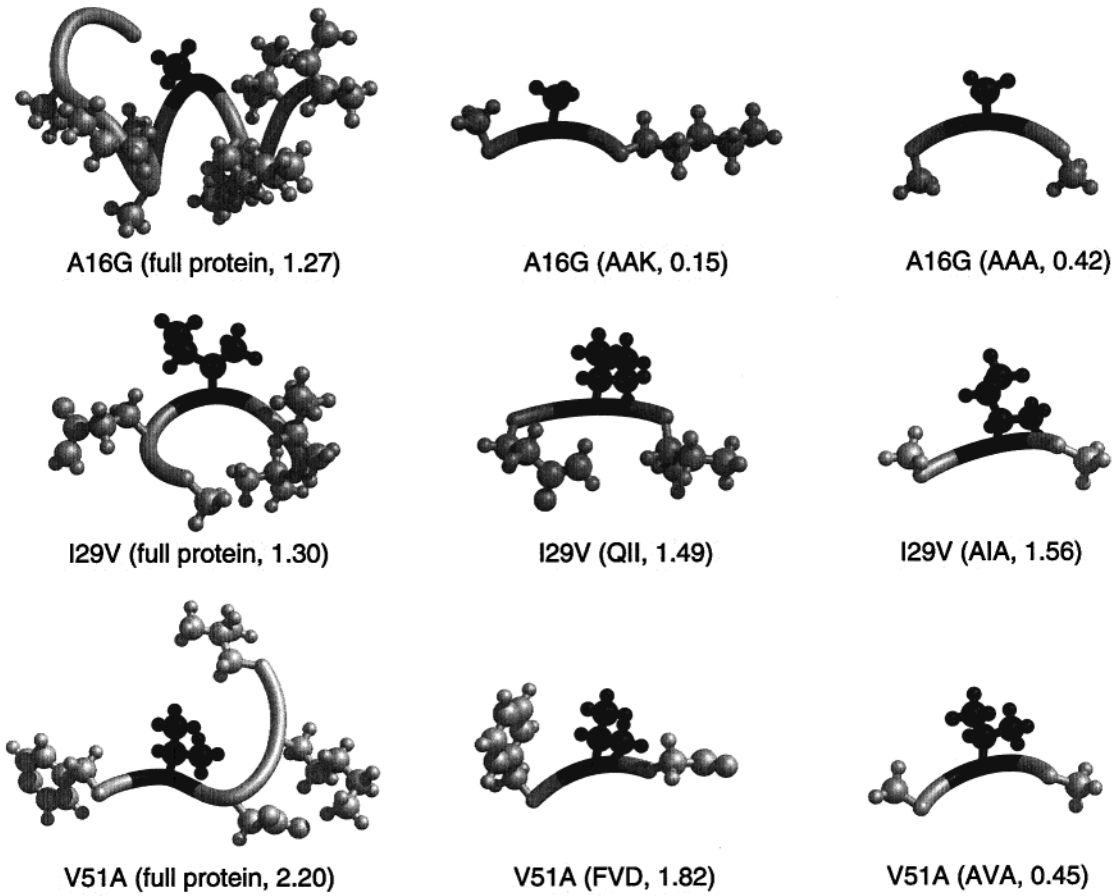

FIGURE 6: Comparison of the local environment of residues A16, I29, and V51 in three different models for the denatured state before mutation. The full protein model is the 3 ns, D1a denatured-state structure, and only the residues within 4 Å of the residue to be mutated are shown. All the residues to be mutated are highlighted in black. The free energy change (ΔG, kcal/mol) for each mutation is also shown in parentheses.

Table 4: Comparison of Calculated ΔG_D and ΔG_{N-D} Values Using Different Models for the Denatured State of CI2[a]

	full protein		tripeptide-1[b]		tripeptide-2[c]		expt
mutation	ΔG_D	ΔG_{N-D}	ΔG_D	ΔG_{N-D}	ΔG_D	ΔG_{N-D}	ΔG_{N-D}
A16G	1.10	1.18	0.15	2.13	0.42	1.86	1.09
V19A	1.09	0.59	1.72	-0.04	0.45	1.23	0.49
I20V	1.42	1.22	0.80	1.84	1.56	1.08	1.30
I29V	1.04	1.41	1.49	0.96	1.56	0.89	1.11
V51A	1.92	2.32	1.82	2.42	0.45	3.79	1.98
I57A	1.80	3.62	2.01	3.41	1.71	3.71	4.29
correlation with expt		0.98		0.84		0.77	

[a] For the full-length protein, the average ΔG_D values from Table 1 were used to calculate ΔG_{N-D}. [b] The protein sequences used for tripeptide-1 were AAK (A16G), LVI (V19A), VIL (I20V), QII (I29V), FVD (V51A), and NIA (I57A). [c] The mutated residues are centered between two alanine residues in this tripeptide model.

studies, the denatured state was assumed to be a random coil, and small, extended peptides were used as surrogates (*32*−*37*). We can now begin to address whether peptides are adequate models for the denatured state. To this end, we performed calculations with two different tripeptide models. One is a simple alanine tripeptide in which the residue to be mutated is surrounded by two alanine residues. The other contains the sequence from the protein with the target residue in the center. Differences between the two tripeptide models are observed (Table 4). The simple Ace-AXA-Nme model and the tripeptide taken from the protein sequence should give similar results, if the peptide is extended with all residues pointing to the solvent. This turns out not to be the case. For example, the AVA to AAA mutation only involves a free energy change of 0.45 kcal/mol while the FVD to FAD mutation is 1.82 kcal/mol (see V51A in Table 4 and Figure 6). The importance of neighboring residues has also been seen in calculations of pentapeptide models by Sugita and Kitao (*36*). They studied five different structures for the same pentapeptide to investigate the conformational influence on the free energy change. The five different starting structures gave different results, and only one starting structure was reasonable when compared with the experimental data.

The results using two different peptides for the same mutation are quite variable: $R = 0.21$ for comparison of the ΔG values for peptides 1 and 2 (Table 4). The values also show little correlation, in general, with the results using the full-length protein models. This discrepancy is due to the complicated local and tertiary interactions in the full-length protein relative to the more limited interactions of simple tripeptides. Our results demonstrate that the simple tripeptide models only reproduce the results using the full-length denatured-state models, and the experimental data, when the environment of that residue in the peptide happens to be similar in the different models (for example, see I29V, Table 4, Figure 6). When the tripeptide model does not agree with the full-length model and experiment, the central residue makes less hydrophobic contacts, leading to a smaller free

energy change upon mutation, as in the case for A16G and V51A (Table 4 and Figure 6). The correlation of $\Delta\Delta G_{N-D}$ values with experiment can best test the reliability of the denatured-state models (Table 4). In our opinion, only the full-length protein models yield satisfactory results.

Why are there such differences between the peptide and full protein models? In the case of peptide 1, which takes its sequence from the protein, the trend with respect to experiment is actually satisfactory ($R = 0.84$), but the absolute differences between the calculated and experimental $\Delta\Delta G_{N-D}$ values are large. The inconsistency of the peptide models to reproduce the values from the full-length protein is because many residues are not fully exposed to the bulk solvent in the full-length denatured state even for this small, unstructured protein. Some of the residues are partially inaccessible to solvent because they are hindered by neighboring residues, even in the absence of secondary and tertiary structures. That the peptide models are less reliable than the models of intact denatured protein is notable because the denatured state of CI2 is very unstructured in both the simulations and experiments (Figure 2) (*7, 8, 29*−*31, 38*−*40*). Furthermore, we would expect the tripeptide model to be even worse for proteins with considerable residual structure in the denatured state. We note that there is experimental precedent for this phenomenon. For example, two helices ($\alpha1$ and $\alpha2$) in barstar contain residual native structure in the denatured state; however, all fragments of these peptides either alone or together are random coils (*41, 42*), suggesting that nonlocal interactions with the rest of the protein are necessary to form or stabilize these helices.

CONCLUSIONS

The results presented here show that short, standard FEP calculations can successfully reproduce experimental free energy changes for hydrophobic deletion mutants of CI2 using transition and denatured states from denaturation simulations. The agreement between our calculated results and experiment is good for both the Φ values and individual $\Delta\Delta G$ values, with correlation coefficients of 0.85 and 0.79, respectively. These values increase to 0.91 and 0.88 if one problematic residue is disregarded. Such agreement suggests that the MD-generated transition- and denatured-state structures are reasonable models. The relationship between the free energy change and the nature of the mutation reveals that the free energy changes are dominated by the extent of hydrophobic contact, as suggested previously by Jackson et al. (*13*) and Matthews and co-workers (*43*−*44*), and the solvent exposure of the mutated residue. The good agreement between the experimental and calculated $\Delta\Delta G_{N-D}$ values for the full-length denatured structures and the inability of tripeptides to consistently reproduce experiment suggest that the use of full-length models of the denatured state is necessary to obtain reliable results. Finally, direct comparison of Φ values from calculated and experimental free energy changes supports our previous transition-state assignments as well as the simulated unfolding pathways and their relevance to refolding.

ACKNOWLEDGMENT

We thank Professors Alan Fersht, Martin Karplus, and Peter Kollman for many helpful discussions and comments on the manuscript. UCSF MidasPlus (*46*) was used to display the structures in Figures 1, 2, 4, 5, and 6.

REFERENCES

1. Fersht, A. R. (1995) *Curr. Opin. Struct. Biol. 5*, 79−84.
2. Matouschek, A., Kellis, J. T,. Jr., Serrano, L., and Fersht, A. R. (1989) *Nature 340*, 122−126.
3. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
4. Serrano, L., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 805−818.
5. Fersht, A. R. (1993) *FEBS Lett. 325*, 5−16.
6. Matouschek, A., Serrano, L., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 819−835.
7. Li, A., and Daggett, V. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10430−10434.
8. Li, A., and Daggett, V. (1996) *J. Mol. Biol. 257*, 412−429.
9. Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1996) *J. Mol. Biol. 257*, 430−440.
10. Li, A., and Daggett, V. (1998) *J. Mol. Biol. 275*, 677−694.
11. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
12. Otzen, D. E., Itzhaki, L. S., elMasry, N. F., Jackson, S. E., and Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10422−10425.
13. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M., and Fersht, A. R. (1993) *Biochemistry 32*, 11259−11269.
14. Jackson, S. E., elMasry, N., and Fersht, A. R. (1993) *Biochemistry 32*, 11270−11278.
15. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
16. Kollman, P. A. (1993) *Chem. Rev. 93*, 2395−2417.
17. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, L. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997) *AMBER 5*; *Computer Program*, University of Califonia, San Fransisco.
18. Harpaz, Y., Elmasry, N., Fersht, A. R., and Henrick, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3−15.
19. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J. Chem. Phys. 79*, 926−935.
20. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. J., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
21. van Gunsteren, W. F., and Berendsen, H. (1977) *J. Mol. Phys. 34*, 1311−1327.
22. Li, A., and Daggett, V. (1995) *Protein Eng. 8*, 1117−1128.
23. Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6−12.
24. Shortle, D., and Abeygunawardana, C. (1993) *Structure 1*, 121−134.
25. Shortle, D. (1996) *FASEB J. 10*, 27−34.
26. Wuthrich, K. (1994) *Curr. Opin. Struct. Biol. 4*, 93−99.
27. Bond, C. J., Wong, K. B., Clarke, J., Fersht, A. R., and Daggett, V. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 13409−13413.
28. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R., and Daggett, V. (2000) *J. Mol. Biol. 296*, 1257−1282.
29. Neira, J. L., Davis, B., Ladurner, A. G., Buckle, A. M., Gay, G. D. P., and Fersht, A. R. (1996) *Fold. Des. 1*, 189−208.
30. Kazmirski, S. L., Wong, K., Freund, S. M. V., Tan, Y., Fersht, A. R., and Daggett, V. (2001) *Proc. Natl. Acad. Sci. U.S.A.* (submitted for publication).
31. Tan, Y. J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 980−992.
32. Tidor, B., and Karplus, M. (1991) *Biochemistry 30*, 3217−3228.
33. Karplus, M., Prevost, M., Tidor, B., and Wodak, S. (1991) *Ciba Found. Symp. 161*, 63−74.

34. Sun, Y.-C., Veenstra, D. L., and Kollman, P. A. (1996) *Protein Eng. 9*, 273−281.
35. Sugita, Y., and Kitao, A. (1998) *Proteins: Struct., Funct., Genet. 30*, 388−400.
36. Sugita, Y., and Kitao, A. (1998) *Biophys. J. 75*, 2178−2187.
37. Yamaotsu, N., Moriguchi, I., and Hirono, S. (1993) *Biochim. Biophys. Acta 1203*, 243−250.
38. Prat Gay, G., Ruiz-Sanz, J., Davis, B., and Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10943−10946.
39. Neira, J. L., Itzhaki, L. S., Ladurner, A. G., Davis, B., de Prat Gay, G., and Fersht, A. R. (1997) *J. Mol. Biol. 268*, 185−197.
40. Tan, Y.-J. (1995) Ph.D. Thesis, Cambridge University, Cambridge, U.K.
41. Wong, K. B., Freund, S. M., and Fersht, A. R. (1996) *J. Mol. Biol. 259*, 805−818.
42. Nölting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 826−830.
43. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) *Science 255*, 178−183.
44. Xu, J., Baase, W. A., Baldwin, E., and Matthews, B. W. (1998) *Protein Sci. 7*, 158−177.
45. Hubbard, S. J., and Thornton, J. M. (1993) *NACCESS, Computer Program*, Department of Biochemistry and Molecular Biology, University College, London.
46. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.

BI0022036