Folding of Circular and Permuted Chymotrypsin Inhibitor 2: Retention of the Folding Nucleus†

Daniel E. Otzen‡ and Alan R. Fersht*

MRC Unit for Protein Function and Design and Cambridge Centre for Protein Engineering, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

Received February 2, 1998; Revised Manuscript Received March 27, 1998

ABSTRACT: The 64-residue chymotrypsin inhibitor 2 (CI2) folds by a two-state nucleation−condensation mechanism, whereby secondary and tertiary structure coalesce concomitantly in the transition state around Ala 16 in the helical N-cap. Permutation of the SH3-domain of α-spectrin apparently shifts its folding nucleus to another region of the protein, suggesting that a protein's transition state may be altered by altering the protein's connectivity. We have characterized the structure of the transition state of a circular and a permuted version of CI2 by a protein engineering study encompassing 11 mutations. Circular CI2 was obtained by the introduction of cysteines at residues 3 and 63 and linking them by disulfide bond formation. Subsequent cyanogen−bromide cleavage of the scissile bond, Met 40−Glu 41, yielded permuted CI2. Circular and permuted CI2 also fold according to a two-state mechanism. Permutation does not affect the folding rate constant, but circularization increases it 7-fold. The transition states of circular and permuted CI2 are essentially unchanged from that of wild-type CI2. Importantly, the folding nucleus around Ala16 is retained. These results complement a previous observation that the transition state for association of two CI2 fragments (residues 1−40 and 41−64, generated by CNBr cleavage) is very similar to the folding transition state of intact CI2. The similarity of rate constants for folding of wild-type and permuted CI2, and their value relative to that for the association of fragments, allows us to estimate the gain in entropy of activation on having the separate fragments linked: 18.3 cal M^{-1} K^{-1}; i.e. an effective molarity of 10^4 M. The contrast between the retention of the folding nucleus on permutation of CI2 and its change for the SH3-domain of α-spectrin probably arises because the latter was cleaved in its folding nucleus whereas cleavage at sites other than 40−41 in CI2 is very destabilizing. Whether or not a folding nucleus can be changed probably depends on the specific protein and its permissivity to permutation.

The N- and C-termini of many proteins are close in space. This allows the construction of artificially permuted proteins, in which the N- and C-termini are linked and new termini are formed elsewhere. Such a permutation occurs naturally, for example, in concanavalin A where it takes place posttranslationally (*1, 2*). On the genetic level, permutations may arise through either gene duplication or excision and reinsertion of the genetic sequence of one part of the protein (*3*). Circular and permuted bovine pancreatic trypsin inhibitor (BPTI), obtained by chemical linkage of the N- and C-terminal residues and subsequent cleavage of the reactive site, both refold correctly (*4, 5*). Genetic permutants of the 250-residue phosphoribosyl-anthranilate-isomerase, made by introducing a linker between the existing termini and transferring the termini elsewhere, also fold correctly and retain wild-type-like stability (*6, 7*). Ten permutations of different proteins have been reported within the last 5 years, namely dihydrofolate reductase (*8*), interleukin 1β (*9*), T4 lysozyme (*10*), the catalytic subunit of the hetero-oligomer aspartate transcarbamoylase (*11, 12*), β-glucanase H from *Bacillus* (*13*), interleukin toxin 4 (*14*), RNase T1 (*15*), the tetrameric glyceraldehyde-3-phosphate dehydrogenase (*16*), dihydrofolate reductase (*17*), and the α-spectrin SH3 domain (*18, 19*). All these "permuteins" retain catalytic activity and significant stability against thermal or denaturant-induced unfolding, showing that the location of the N- and C-termini of both monomeric and oligomeric proteins have remarkably little effect on enzymatic activity and overall structure. Interestingly, permutation of the SH3 domain appears to change the folding pathway despite minimal effect on the native structure (*18, 19*). Whether or not there is a unique folding nucleus is a matter of some dispute according to various theoretical analyses (*20*−*22*).

The 64-residue chymotrypsin inhibitor 2 (CI2) is suitable for permutation, since the two termini are in close proximity. One face of a six-stranded β-sheet packs against a single α-helix to form the hydrophobic core, while an extended loop, which binds to subtilisin BPN′ and chymotrypsin, projects from the other face (*23, 24*). The two termini are at the ends of the first and last β-strand. CI2 folds according to a classical two-state model involving just the unfolded and folded state under both equilibrium and nonequilibrium conditions (*25*). The folding of CI2 is known in great detail from a protein engineering approach (*26, 27*) complemented by molecular dynamics simulations (*28*). The transition state of CI2 is an expanded version of the native state, in which only the residues near the C- and N-termini are completely

† D.E.O. was supported by a predoctoral fellowship from the Danish Natural Science Research Council.

* To whom correspondence should be addressed.

‡ Present address: Department of Biochemistry, Chemistry Centre, University of Lund, P.O. Box 124, S − 22100 Lund.

S0006-2960(98)00250-5 CCC: $15.00
Published on Web 05/14/1998

unfolded, while the interactions among side chains in the rest of the protein are partially formed (*26*, *27*). Folding is driven by the condensation of secondary and tertiary structure around Ala 16 in the N-cap of the α-helix (the folding nucleus). We know from the association of fragments of CI2 cleaved at Met 40 that the folding nucleus is unchanged on cleavage at that point, as is the structure of the complex of the fragments in solution or the crystal (apart from the cleaved bond) (*29*).

Since the terminal Gly 64 forms a vital salt bridge with Arg 46, genetic permutation of CI2 is not feasible. Instead, we have permuted CI2 by replacing residues 3 and 63 with cysteine, linking them with a disulfide bond and cleaving the scissile bond between Met 40 and Glu 41 in the reactive site loop with cyanogen bromide (*30*). In contrast to permutation at the genetic level, this allows us to examine the transition states of both the circular and permuted forms. Eleven residues have been mutated to Ala or Gly in different regions, namely the N-terminus (Lys 2), the β-sheet (Val 53, Ala 58, and Val 60), the interface between the hydrophobic core and the N-cap of the α-helix (Ala 16), a hydrophobic patch (Leu 32, Val 38, and Phe 50), the reactive site loop (Tyr 42 and Arg 46), and a turn connecting the first β-strand with the α-helix (Pro 6). Ala 16 probes the formation of the folding nucleus, while Lys 2 and Val 60 in the region around the disulfide bond investigate whether the transition state of folding (which, in the absence of disulfide bonds, is completely unfolded in this part of the protein) is locally perturbed by the introduction of the cross-link. The results show that the transition state of folding is broadly unperturbed by the change in connectivity, though there are subtle changes.

MATERIALS AND METHODS

Materials. All materials used were as described (*31*, *32*). Guanidinium thiocyanate (GdmSCN) grade SigmaUltra (>99%) was from Sigma.

Preparation of Circular CI2. Mutants were prepared, purified, and confirmed as described (*31*). Circularization via oxidation of the two Cys mutants occurred spontaneously during purification. Ellman assays (*33*) did not reveal free dithiol groups in the presence or absence of 6 M guanidinium chloride (GdmCl), and gel electrophoresis in the absence of reducing agents ruled out the existence of dimers or higher-order structures.

Preparation of Permuted CI2 from Circular CI2. Circular CI2 was cleaved by cyanogen bromide (CNBr) by incubating 2.4 mg/mL protein for 30 min at pH 2.3 (pH adjusted by HCl), adding CNBr to a final concentration of 2 CNBr:1 protein (w/w), incubating at room temperature for 6−7 h and lyophilizing. Cleaved protein was separated from uncleaved protein on a Superdex 75 HR 10/30 gel filtration column in 50 mM MES, pH 6.25, and high concentrations of GdmCl (3−6 M). Under these conditions, circular CI2 remains folded while permuted CI2 is predominantly unfolded and therefore has a larger hydrodynamic volume, leading to a smaller elution volume. A suitable concentration of GdmCl was empirically found to be $2[\text{GdmSCN}]^{50\%} - 1.3$ M, where $[\text{GdmSCN}]^{50\%}$ is the concentration of guanidinium thiocyanate (GdmSCN) at which the circular mutant is 50% unfolded. Permuted CI2 retained the molecular weight of the protein in the absence, but not the presence, of reducing agents as judged by gel electrophoresis. Protein was dialyzed extensively against water and stored at −80 °C.

Equilibrium Denaturation. Equilibrium denaturation in GdmCl or GdmSCN monitored by fluorescence spectroscopy was carried out essentially as described (*31*). GdmSCN was needed to unfold circular CI2 mutants. CI2 (1.5 μM) in 50 mM MES, pH 6.25, and different concentrations of denaturant were preincubated for at least 1 h at 25 °C before measuring the fluorescence intensity. To reduce circular CI2, 180 mM DTT was preincubated with the protein stock solution for 30 min at room temperature to give a final concentration of 20 mM DTT after mixing with denaturant. Circular CI2 was incubated with 9 mM oxidized DTT for 30 min at room temperature before mixing with denaturant. The fluorescence data are analyzed as described (*34*), using an average value ($\langle m \rangle$) of m_{U-F}, the sensitivity of the free energy of unfolding to denaturant. The value of $\langle m \rangle$ for unpermuted mutants (using GdmCl) is 1.92 ± 0.03 kcal mol^{-1} M^{-1}. For circular mutants, $\langle m \rangle$ is 2.53 ± 0.05 kcal mol^{-1} M^{-1} (in GdmSCN) and, for permuted mutants, $\langle m \rangle$ is 1.48 ± 0.04 kcal mol^{-1} M^{-1} (in GdmCl). These mean values allow calculation of $\Delta\Delta G_{U-F}^{[D]50\%}$ with low standard error (*34*).

Unfolding and Refolding Experiments. Kinetic experiments were carried out at 25 °C as described (*27*, *35*). DTT (110 μM) was included with TC3/VC63 to give a final concentration of 10 μM after mixing with denaturant solution. For unfolding studies, 1 vol of protein solution (about 50 μM) in 50 mM MES, pH 6.25, was rapidly mixed with 10 vol of denaturant (in 50 mM MES, pH 6.25) to different final denaturant concentrations. Refolding studies were carried out by mixing 1 vol of denatured protein (unfolded in either 6.5 M GdmCl or 4.4 M GdmSCN) with 10 vol of refolding buffer containing a low concentration of denaturant in 50 mM MES, pH 6.25. The lowest final denaturant concentration obtained this way was 0.591 M GdmCl or 0.4 M GdmSCN. pH-jump refolding studies at low denaturant concentrations were not performed, since circular and permuted CI2 are not unfolded at low pH. Refolding of CI2 is multiphasic and includes the major fast refolding phase (with refolding rate k_f), which accounts for about 70% of the total amplitude change on refolding, as well as slower phases associated with proline isomerization (*36*). Only the fast phase is important for this discussion. Refolding curves of TC3/VC63 were fitted to a double exponential curve with baseline drift, while refolding of circular and permuted CI2 was adequately fitted to a single curve (no significant improvement in the residuals was observed when double exponentials were used, and the amplitudes and rates of the two phases derived from the double exponential fit were too close to each other to be easily distinguished). This does not rule out proline isomerization in the unfolded state of circular and permuted CI2, though it suggests that the isomerization is slowed by the disulfide bond.

Analysis of Kinetic Data. The logarithms of the unfolding constant k_u and the refolding constant k_f are linearly related to denaturant concentration (*35*). Rates of refolding and unfolding in water may be calculated by extrapolation, except for the refolding rates of wild-type CI2 and TC3/VC63, which can be measured directly by pH-jump refolding. A linear relationship between the unfolding and refolding

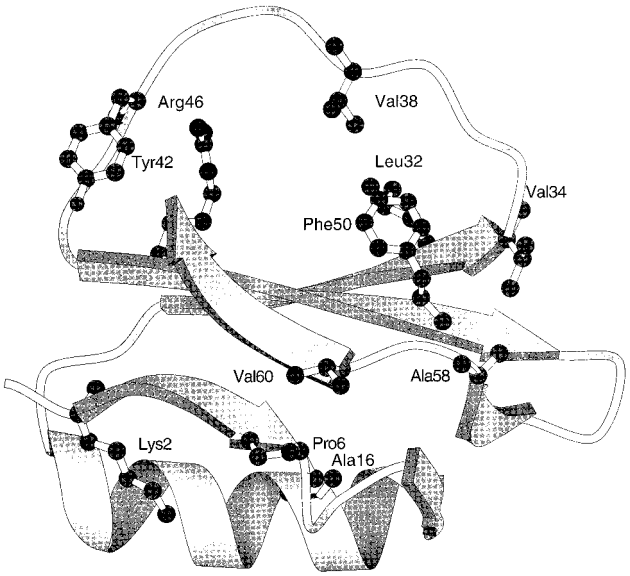

FIGURE 1: Structure of CI2, showing positions of side chains mutated in this study. The structure was generated using the program Molscript (*46*).

constants and the concentrations of GdmSCN is also observed over the measured concentration range (0.5−2.6 M). The refolding of circular CI2 is too fast to be measured below 1.5 M GdmSCN. To reduce the error of extrapolation, refolding rates from experiments with GdmSCN were therefore calculated at 1.5 M GdmSCN.

The rate constants of unfolding and refolding measure the difference in energy between the transition state and the folded and unfolded states, respectively. The difference in activation energy of unfolding and refolding of the various mutants relative to their respective wild-types ($\Delta\Delta G_{\ddagger-F}$ and $\Delta\Delta G_{\ddagger-U}$) are calculated as described (*35*) at 4 and 1.5 M GdmSCN (circular mutants), and 6 and 0 M GdmCl (permuted mutants).

RESULTS

Disulfide Bond Description. The position of the disulfide bond in CI2 was selected using the following criteria. (1) It should link two residues of the protein that were as far apart as possible in sequence. (2) The C^{β} atoms of the two residues should be within 3.5−4.5 Å of each other (*37*, *38*). (3) The structural and thermodynamic consequences of introducing a Cys residue should be small, i.e., Cys should replace residues with side chains of the same size and avoid removing stabilizing interactions with other residues. On this basis, Thr 3 and Val 63 were selected. These residues are in strand 1 and strand 6 of the β-sheet; the carboxyl oxygen of Thr 3 forms a hydrogen bond to the amide nitrogen of Val 63 (Figure 1). The $C^{\beta}-C^{\beta}$ distance is 4.2 Å. The replacement of Thr and Val by Cys is a virtually isosteric mutation.

Stability Data. Circular and permuted mutants of CI2-fold by a two-state mechanism under equilibrium conditions. There is a significant difference between the values of m_{U-F} for TC3/VC63, oxTC3/VC63 (the circular form of TC3/VC63), and permTC3/VC63 (oxTC3/VC63 cleaved at Met 40), which means that their stabilities relative to each other ($\Delta\Delta G_{U-F}$) varies with denaturant concentration (Table 1). $\Delta\Delta G_{U-F}$ can be calculated at any concentration of denaturant (D) by applying the following equation (*34*):

$$\Delta\Delta G_{U-F}{}^{D} = m_{U-F}([D]^{50\%} - [D]) - m_{U-F}'([D']^{50\%} - [D]) \quad (1)$$

$[D]^{50\%}$ is the denaturant concentration where 50% of the protein population is denatured. The prime denotes the variant form of CI2 (circular or permuted). m_{U-F} is a measure of the increase in exposure of protein surface to solvent on unfolding (*39*). Therefore, the lower values of m_{U-F} for oxTC3/VC63 and permTC3/VC63 could be interpreted as evidence of residual structure in the unfolded states of the two disulfide-bond variants because of the conformational restrictions imposed by the disulfide bond. However, the lower values of m_{U-F} could also arise because the conformational restrictions of the disulfide bond in both circular and permuted CI2 decrease the binding of denaturant to the unfolded protein. The stabilizing effect of a disulfide bond is then a balance between the loss of conformational freedom in the unfolded state and the reduction in the number of denaturant molecules that can bind to this state.

Stability data for the different CI2 mutants are shown in Table 2. A plot of $\Delta\Delta G_{U-F}{}^{[D]^{50\%}}$ for unpermuted mutants versus the corresponding values of $\Delta\Delta G_{U-F}{}^{[D]^{50\%}}$ for circularized mutants (Figure 2A) gives a good linear correlation [the fitted line has an intercept of zero (0.04 ± 0.12) and slope of 0.85 ± 0.06]. This proportionality indicates that the effect of the disulfide bond on the stability of CI2 is uniformly distributed throughout the protein. The corresponding plot between unpermuted mutants and permuted mutants (Figure 2B) is more scattered. Only a subset of the mutants is destabilized to the same extent in TC3/VC63 and permuted CI2. These mutations are all in regions of CI2 that are distant from the cleaved bond, namely KA2, PA6, AG16, AG58, and VA60; two of them (KA2 and VA60) are close to the disulfide bond. Even in the absence of the disulfide bond, the backbone and side chains of the termini of TC3/VC63 are constrained by hydrogen bonds and packing interactions, so the disulfide bond is not expected to increase the rigidity of this region. In contrast, all the mutants that lie below the indicated line, namely LA32, VG34, VA38, YA42, RA46, and FA50, are part of, or project into, the reactive site loop. This region is more flexible because of the cleavage between Met 40 and Glu 41, and therefore, these side chains do not make as large contributions to protein stability in permuted CI2 via packing interactions as in unpermuted CI2. A similar segregation is observed for kinetic data.

Kinetic Data. All the CI2 mutants in this study fold by a two-state mechanism as defined by the usual criteria (V-shaped plot of ln k_{obs} versus denaturant concentration and correspondence between free energies, destabilization energies and m-values from equilibrium and kinetic experiments) (*27*). The m_{k_u}-values of oxTC3/VC63 and permTC3/VC63 are consistently lower than those of TC3/VC63 in GdmSCN and GdmCl (Table 3). This means that the change in activation energy of unfolding compared to TC3/VC63 is very denaturant dependent. In water, oxTC3/VC83 unfolds 11 times slower than TC3/VC63; this increases to 161 in 3 M GdmSCN. The protein engineering data (below) reveal that the terminal residues are still unfolded in the transition

Table 1: Equilibrium Denaturation Parameters[a]

protein (denaturant)	$[D]^{50\%}$ (M)[b]	m_{U-F}[c] (kcal mol^{-1} (kcal mol^{-1})	$\Delta G_{U-F}^{H_2O}$ (kcal mol^{-1} M^{-1})	$\Delta\Delta G_{U-F}^{H_2O}$ [d] (relative to TC3/VC63) (kcal mol^{-1})	$\Delta\Delta G_{U-F}^{[D]50\%}$ [d] (relative to TC3/VC63) (kcal mol^{-1})
wild-type (GdmSCN)	2.10 ± 0.01	3.45 ± 0.11	7.22 ± 0.24		
TC3/VC63[g] (GdmSCN)	1.80 ± 0.01	3.81 ± 0.16	6.87 ± 0.29		
oxTC3/VC63[h] (GdmSCN)	3.32 ± 0.01	2.41 ± 0.09	8.00 ± 0.31	−1.12 ± 0.42	−4.73 ± 0.16[e]
wild-type (GdmCl)	4.00 ± 0.01	1.90 ± 0.03	7.60 ± 0.12		
TC3/VC63 (GdmCl)	3.62 ± 0.02	1.90 ± 0.06	6.86 ± 0.22		
permTC3/VC63[i] (GdmCl)	4.57 ± 0.02	1.51 ± 0.05	6.91 ± 0.24	−0.05 ± 0.37	−1.62 ± 0.06[f]

[a] All experiments were carried out at 25 °C and pH 6.25. [b] The concentration at which 50% of the protein population is denatured. [c] The constant from the equation $\Delta G_{U-F}^{D} = \Delta G_{U-F}^{H_2O} + m_{U-F}D$, where $\Delta G_{U-F}^{H_2O}$ and ΔG_{U-F}^{D} are the free energies of folding in water and at a given denaturant concentration. [d] Calculated using eq 1. [e] Calculated at 2.56 M GdmSCN, midway between the $[GdmSCN]^{50\%}$ of TC3/VC63 and oxTC3/VC63. [f] Calculated at 4.10M GdmCl, midway between the $[GdmCl]^{50\%}$ of TC3/VC63 and permTC3/VC63. [g] In 20 mM DTT. [h] In 1 mM oxidized DTT. [i] Permuted TC3/VC63 (intact disulfide bond but cleaved at Met 40).

Table 2: Equilibrium Denaturation Parameters for CI2 Mutants[a]

mutant	circularized mutants (GdmSCN $[GdmSCN]^{50\%}$ (M)	permuted mutants (GdmCl) $[GdmCl]^{50\%}$ (M)	unpermuted[b] mutants (GdmCl) $[GdmCl]^{50\%}$ (M)
ref[c]	3.32 ± 0.01	4.57 ± 0.02	4.00 ± 0.01
KA2	3.19 ± 0.01	4.12 ± 0.02	3.72 ± 0.01
PA6	2.61 ± 0.01	3.51 ± 0.02	3.19 ± 0.05
AG16	2.87 ± 0.01	3.78 ± 0.01	3.44 ± 0.02
LA32	2.53 ± 0.01	3.61 ± 0.01	2.78 ± 0.02
VG34	2.53 ± 0.02	3.83 ± 0.01	2.75 ± 0.01
VA38	2.83 ± 0.01	4.48 ± 0.02	3.24 ± 0.02
YA42	2.79 ± 0.01	4.08 ± 0.03	3.01 ± 0.01
RA46	2.34 ± 0.01	3.20 ± 0.06	2.49 ± 0.02
FA50	1.99 ± 0.02	2.99 ± 0.02	2.02 ± 0.03
AG58	2.63 ± 0.01	3.42 ± 0.02	3.03 ± 0.04
VA60	2.79 ± 0.01	3.78 ± 0.03	3.22 ± 0.03

[a] All experiments were carried out at pH 6.25 and 25 °C. Data fitted satisfactorily to a two-state model. [b] Itzhaki et al. (*27*). [c] oxTC3/VC63 for circularized mutants, permTC3/VC63 for permuted mutants, and wild-type CI2 for unpermuted mutants.

states of circular and permuted CI2, but since the termini are in close proximity at all stages of the folding, they contribute less to the overall increase in solvent-accessible surface area in going from the native state to the transition state than they would in wild-type CI2. This leads to a lower value of m_{k_u}.

However, the dependence of ln k_f on denaturant concentration (m_{k_f}) is very similar for the different variants of CI2 (Table 4), in other words the increase in exposure of protein surface to solvent on going from the unfolded to the transition state is not affected by the disulfide bond (because the terminal regions remain just as unfolded in the transition state as in the unfolded state, irrespective of the disulfide bond). This makes the change in energy of activation of unfolding compared to TC3/VC63 virtually independent of denaturant concentration. At 1.5 M GdmSCN, which is in the concentration range where the rate can be measured directly, oxTC3/VC63 refolds significantly faster than TC3/VC63 (12.8 vs 2.5 s^{-1}, respectively). This extrapolates to 440 vs 100 s^{-1} in water. permTC3/VC63 refolds at a rate very similar to TC3/VC63. Thus, the structural restriction introduced by the disulfide bond in the unfolded state reduces the conformational search for the native fold.

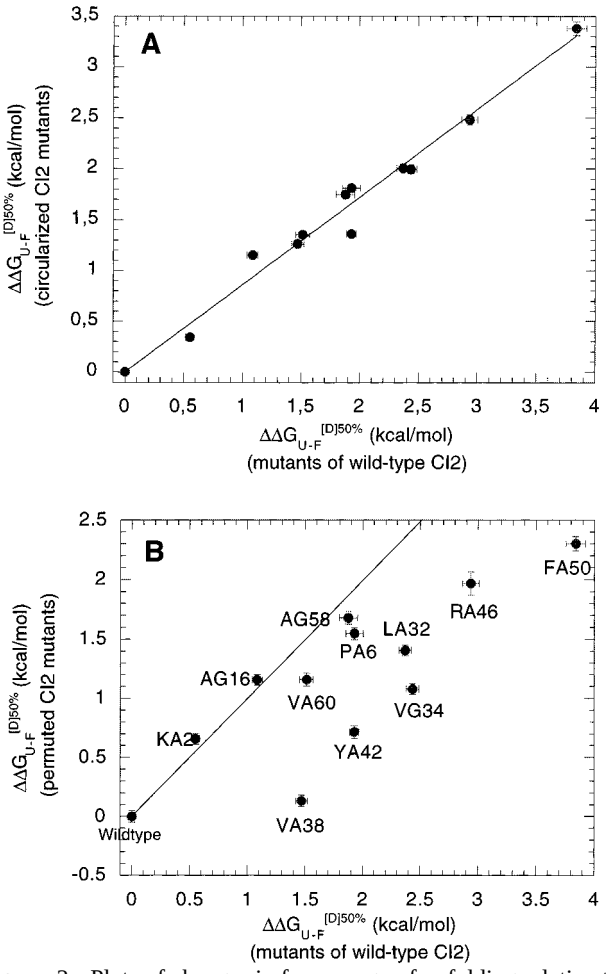

FIGURE 2: Plots of changes in free energy of unfolding relative to wild-type ($\Delta\Delta G_{U-F}^{[D]50\%}$) for mutants without disulfide bonds versus $\Delta\Delta G_{U-F}^{[D]50\%}$ for (A) circular mutants and (B) permuted mutants. The straight line in panel A is the best linear fit, whereas the line in panel B (which has a slope of one and an intercept of zero) is purely illustrative.

Protein Engineering Analysis. Mutations of individual side chains in permuted CI2 enable us to characterize the transition state of permuted CI2 and compare it with that of unpermuted CI2. The folding nucleus of unpermuted CI2's transition state principally involves three residues, namely

Table 3: Unfolding Parameters[a]

protein (denaturant)	m_{k_u} (M^{-1})[b]	$k_u^{H_2O}$ (s^{-1}) ($\times 10^{-3}$)	k_u^D (s^{-1})	$\Delta\Delta G_{\ddagger-F}^{H_2O}$ [c] (relative to TC3/VC63) (kcal mol^{-1})	$\Delta\Delta G_{\ddagger-F}^{D}$ [e] (relative to TC3/VC63) (kcal mol^{-1})
wild-type (GdmSCN)	2.39 ± 0.04	2.82 ± 0.34	3.63 ± 0.04[d]		
TC3/VC63 (GdmSCN)	2.48 ± 0.06	7.91 ± 1.19	13.3 ± 0.80[d]		
oxTC3/VC63 (GdmSCN)	1.57 ± 0.02	0.73 ± 0.09	0.08 ± 0.01[d]	-1.41 ± 0.11	-3.01 ± 0.05[d]
wild-type (GdmCl)	1.31 ± 0.01	0.12 ± 0.01	0.27 ± 0.01[e]		
TC3/VC63 (GdmCl)	1.36 ± 0.01	0.36 ± 0.03	1.26 ± 0.01[e]		
permTC3/VC63 (GdmCl)	1.07 ± 0.03	0.18 ± 0.03	0.11 ± 0.001[e]	-0.40 ± 0.11	-1.44 ± 0.01[e]

[a] All experiments were carried out at 25 °C and pH 6.25. [b] The constant from the equation $\ln k_u^D = \ln k_u^{H_2O} + m_{k_u} D$, where k_u^D and $k_u^{H_2O}$ are the unfolding constants at a given denaturant concentration and in water. [c] Calculated from the equation $\Delta\Delta G_{\ddagger-F}^{D} = -RT \ln(k_u^{D,wild-type}/k_u^{D,mutant})$. [d] At 3 M GdmSCN. [e] At 6 M GdmCl.

Table 4: Refolding Parameters[a]

protein (denaturant)	m_{k_f} (M^{-1})[b]	k_f (s^{-1})	$\Delta\Delta G_{\ddagger-F}$ (relative to TC3/VC63) (kcal mol^{-1})[c]
wild-type (GdmSCN)	-2.47 ± 0.04	2.39 ± 0.05[d]	
TC3/VC63(GdmSCN)	-2.47 ± 0.07	2.51 ± 0.10[d]	
oxTC3/VC63 (GdmSCN)	-2.36 ± 0.05	12.81 ± 0.64[d]	-0.96 ± 0.04[d]
wild-type (GdmCl)	-1.84 ± 0.03	56.8 ± 2.3[e]	
TC3/VC63 (GdmCl)	-2.00 ± 0.01	58.0 ± 1.2[e]	
permTC3/VC63 (GdmCl)	-1.72 ± 0.05	66.7 ± 6.7[e]	-0.08 ± 0.06[e]

[a] All experiments were carried out at 25 °C and pH 6.25. [b] The constant from the equation $\ln k_f^D = \ln k_f^{H_2O} + m_{k_f} D$, where k_f^D and $k_f^{H_2O}$ are the refolding constants at a given denaturant concentration and in water. [c] Calculated from the equation $\Delta\Delta G_{\ddagger-U}^{D} = -RT \ln(k_f^{D,wild-type}/k_f^{D,mutant})$. [d] At 1.5 M GdmSCN. [e] At 0 M GdmCl.

Ala 16, Leu 49, and Ile 57 (*27*). These three residues were therefore obvious candidates for mutagenesis in permuted CI2. Unfortunately, the mutants LA49 and IA57, as well as IV20, IA29 (in which side chains in the hydrophobic environment surrounding the folding nucleus are truncated), and SA12 (the N-cap of the α-helix), failed to express at sufficient levels for further characterization. Our analysis of the transition state of unpermuted CI2 is therefore restricted to the 11 mutants in this study.

The kinetic data in Tables 5 and 6 are combined with equilibrium data in Table 2 to calculate Φ_F-values for individual side chains (Figure 3). A Φ_F-value of 0 indicates that the side chain is as unfolded in the transition state as in the unfolded state, while a value of 1 means that the side chain is as well-formed as in the native state (*40*). In unpermuted CI2, Lys 2 shows unusual nonnative interactions in the transition state (*27*): the mutant KA2 refolds faster than wild-type, even though the mutation destabilizes CI2 toward unfolding under equilibrium conditions. This nonnative behavior is confirmed by data from unfolding experiments. The nonnative interactions are very specific and only involve one other side chain, namely Asp 23, since this is the only other side chain in the vicinity of Lys 2 which also exhibits a negative Φ_F-value, and a double-mutant cycle analysis of the coupling between these two residues reveals a negative Φ_F^{int} (*27*). Lys 2's negative Φ_F-value is retained and even slightly increased in the circular form, but changes and becomes positive in the permuted form. We interpret this as evidence for an increase in nativelike interactions around Lys 2 at the expense of nonnative interactions when CI2 is permuted. It cannot be ruled out that nonnative interactions are still present to some degree. This could for example be the case if permuted CI2 folds by several parallel pathways, in which nonnative interactions are present in some but not all pathways. However, given that unpermuted CI2

Table 5: Unfolding Parameters for Circular and Permuted Mutants of CI2[a]

mutant	circularized mutants (GdmSCN) k_u^{4M} (s^{-1})[c]	permuted mutants (GdmCl) k_u^{6M} (s^{-1})[c]	unpermuted mutants (GdmCl)[b] k_u^{4M} (s^{-1})[c]
ref	0.39 ± 0.01	0.11 ± 0.001	0.020 ± 0.001
KA2	0.99 ± 0.02	0.26 ± 0.005	0.066 ± 0.002
PA6	7.77 ± 0.23	1.35 ± 0.03	0.20 ± 0.03
AG16	0.38 ± 0.01	0.12 ± 0.003	0.014 ± 0.001
LA32	5.53 ± 0.06	0.82 ± 0.02	0.262 ± 0.016
VG34	2.51 ± 0.05	0.54 ± 0.005	0.188 ± 0.002
VA38	1.79 ± 0.05	0.12 ± 0.002	0.088 ± 0.001
YA42	3.86 ± 0.08	0.364 ± 0.01	0.368 ± 0.004
RA46	29.1 ± 3.5	1.336 ± 0.05	3.254 ± 0.10
FA50	11.4 ± 0.23	1.954 ± 0.02	0.72 ± 0.007
AG58	2.51 ± 0.05	0.97 ± 0.01	0.169 ± 0.003
VA60	5.81 ± 0.06	0.835 ± 0.01	0.19 ± 0.006

[a] All experiments were carried out at pH 6.25 and 25 °C. [b] Itzhaki et al. (*27*). [c] Defined as in Table 3.

Table 6: Refolding Parameters for Circular, Permuted and Unpermuted Mutants of CI2[a]

mutant	circularized mutant (GdmSCN) $k_{ff}^{1.5M}$ (s^{-1})[c]	permuted mutants (GdmCl) $k_f^{H_2O}$ (s^{-1})[c]	unpermuted mutants (GdmCl)[b] $k_f^{H_2O}$ (s^{-1})[c]
ref	12.8 ± 0.5	66.7 ± 0.67	56.3 ± 2.3
KA2	15.3 ± 0.5	56.3 ± 3.9	67.4 ± 1.3
PA6	20.1 ± 0.8	66.0 ± 7.2	47.5 ± 0.5
AG16	2.00 ± 0.04	7.61 ± 0.15	8.09 ± 0.16
LA32	7.85 ± 0.08	37.3 ± 1.9	26.8 ± 3.0
VG34	5.87 ± 0.12	40.4 ± 2.8	28.8 ± 1.2
VA38	10.8 ± 0.2	47.5 ± 1.4	41.7 ± 1.3
YA42	15.3 ± 0.05	44.3 ± 1.3	43.4 ± 0.9
RA46	41.7 ± 2.5	48.4 ± 3.9	154.5 ± 3.1
FA50	3.53 ± 0.14	21.1 ± 1.9	8.00 ± 0.9
AG58	8.59 ± 0.09	39.6 ± 4.0	40.9 ± 1.3
VA60	20.3 ± 0.41	47.5 ± 2.4	60.9 ± 2.4

[a] All experiments were carried out at pH 6.25 and 25 °C. [b] Itzhaki et al. (*27*). [c] Defined as in Table 4.

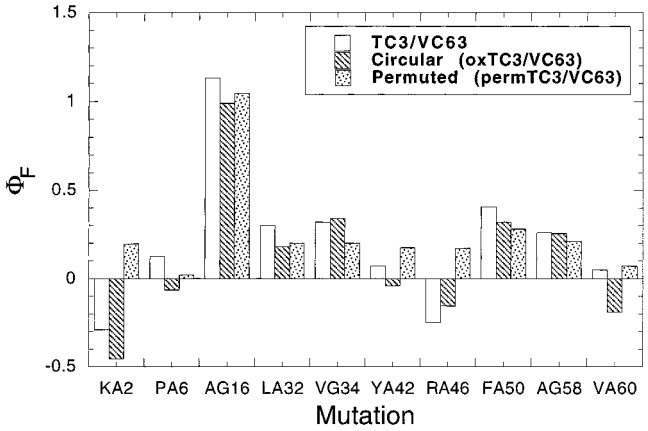

FIGURE 3: Φ-Values for unpermuted mutants, circular mutants, and permuted mutants. Φ-Values are averages of Φ_F (from refolding data) and $1 - \Phi_U$ (from unfolding data).

folds by a single pathway (*41*), this is unlikely. It is more probable that some, but not all, of the nonnative interactions probed by the Lys → Ala mutation (involving various atomic contacts with the side chain of Asp 23) become nativelike. Similarly for Pro 6, the nonnative interactions that are generated by the disulfide bonds are removed by the permutation. Two other mutations probe the structure of the transition state in the immediate vicinity of the disulfide bond, namely AG58 and VA60. The Φ-values of AG58 are not changed by circularization and permutation, but circularization appears to force Val 60 into nonnative interactions (as indicated both by Φ_F and $1 - \Phi_U$), which are removed by permutation. Thus, by making the termini more rigid in the transition state, the disulfide bond enhances preexisting nonnative conditions. Conversely, cleavage of the protein in the reactive loop appears to transmit a relaxation of structure to the termini in the transition state which subtly modifies the structure of the transition state.

The mutation AG16 is our only probe of the core of CI2. Its side chain, which is in the very center of the core, is unique among the 100 mutants of CI2 examined in this way in being completely structured in the transition state of unpermuted CI2 (*27*). Neighboring core residues are folded to a lesser extent. Significantly, the Φ-values of AG16 are completely unaffected by circularization and permutation.

Φ-Values of VG34 are unchanged by circularization but decrease with permutation, suggesting that the cleavage of the nearby scissile residue generally loosens the structure of this part of CI2 in the transition state. The highly solvent-exposed side chain of Tyr 42 is essentially unfolded in unpermuted CI2 and also in circular and permuted CI2. The three residues Leu 32, Val 38, and Phe 50 constitute a local patch of hydrophobic side chains at one end of the extended loop, which we call the minicore. The minicore is partially formed in the transition state of unpermuted CI2. The same level of partial formation is also observed in circular and permuted CI2. Φ-Values for Val 38 in permuted CI2 could not be determined accurately, because the mutation VA38 only destabilizes permuted CI2 to a small extent (Table 1).

Arg 46 shows a very unusual level of nonnative interactions in the transition state of unpermuted CI2 with Φ-values for folding and unfolding of around −0.4. This is attributed to the interactions between its guanidine group and the guanidine group of the proximal Arg 48 (A. G. Ladurner, D. E. Otzen, & A. R. Fersht, unpublished results). We still observe this nonnative behavior in circular CI2, but permutation converts the interaction to a nativelike one, i.e., the Φ-values become positive and fractional. Thus, the structural constraints that caused the nonnative interactions have been removed by permutation.

DISCUSSION

The Folding Nucleus of CI2 Is Retained Despite Fragmentation and Permutation. The disulfide bond engineered into CI2 in this study can be regarded as a nativelike tertiary interaction that is present in the unfolded protein and remains covalently fixed throughout the folding process. A priori, this premature fixing of a nativelike interaction, particularly one that links such distant residues, might be expected to affect a protein's folding mechanism by, e.g., changing the order or number of individual steps or favoring particular pathways. For instance, the folding pathway of the three-disulfide bovine pancreatic inhibitor from the reduced state consists of a series of intermediates with different disulfide bonds, some of them nonnative (*42*). When the two termini are chemically cross-linked by a peptide bond, the major effect is to stabilize a late intermediate lacking the disulfide nearest the termini, while the native state and the early folding intermediates are not stabilized (*5*). In this case, the chemical cross-link appears to have introduced considerable strain into the cross-linked region and skewed the population of intermediates rather than acting as a probe of the folding pathway, as it does in barnase (*34*). However, this complication is not observed with CI2. Unpermuted CI2 folds in a single step with a transition state in which the different regions of CI2 with fractional Φ-values are genuinely only partially folded; they do not consist of a mixture of fully formed or fully unfolded structures which could arise from parallel pathways (*41*). This folding mechanism appears to be robust toward shuffling of termini, since the folding nucleus around Ala 16 retains its nativelike character in the transition state. In fact, Ala 16 is the residue least affected by circularization and permutation. We would have liked to include data on Leu 49 and Ile 57, which also form part of the folding nucleus, as well as Ser 13, Ile 20, and Ile 29, which also show high Φ_F-values in unpermuted CI2 (*27*). Unfortunately, these mutants, when constructed, did not express at detectable levels. However, Ala 16 is the best probe of the folding nucleus since its C^{β}-atom is completely buried in native CI2 and only makes contacts with side chains outside the α-helix (Leu 8, Leu 49, and Ile 57). Therefore, its Φ_F-value of 1 is likely to represent the retention of the entire folding nucleus. Further, the apparent retention of the original folding nucleus in circular and permuted CI2 is supported by studies on the association and folding of equivalent fragments of CI2: the transition state for the association and folding of the fragments 1−40 and 41−64 has been shown to be similar to that of intact CI2 (*29*). In addition to Ala 16, the side chains of Ser 12, Ile 20, Ile 29, and Ile 57 showed the same high Φ_F-values in the transition state of fragment association and folding as in the transition state of unpermuted CI2 (Leu 49 was not analyzed in this system). Clearly the nucleus remains specific under these different conditions. In addition, the side chains near the termini, which are completely unstructured in unpermuted CI2, remain unfolded in the transition states of circular and

permuted CI2, even though the disulfide bond brings them into proximity with residues with which they interact in the native state.

A contrast is provided by the 57-residue α-spectrin SH3 domain, which is also proposed to fold according to a nucleation-condensation mechanism without intermediates (*18*, *19*). Permutations by cleavage in the proposed nucleus (the distal loop and the RT loop) appear to shift the nucleus toward the n-Src loop as well as the N-terminus. This conclusion is based principally upon the mutation Val → Ala44, which has a much lower Φ-value in the N47−D48 permutation than in the original protein. The authors propose that the folding nucleus is not necessarily specific.

The nonuniqueness of the folding nucleus is also proposed from a model off-lattice study by Guo and Thirumalai (*22*), where different conformations in the unfolded state of a 46-mer leads to different nucleation sites and folding rates. Conversely, it has been suggested that folding nuclei are conserved (*20, 21*). A possible explanation for the apparent nuclear shift in SH3 is that the apparent Φ-value of Val 44 in the permuted protein is lowered by nonnative interactions in the unfolded state (*21*) rather than by a loss of interactions in the transition state: the neighboring Lys 43 actually shows a negative Φ-value in permuted SH3. More mutations in the vicinity of Val 44 would be required to clarify this.

A real difference between CI2 and SH3 is that cleavage of CI2 at sites other than the 40−41 bond greatly destabilizes the protein, so much so that the only complex between complementary peptides of CI2 that has been be detected is the 1−40 peptide with the 41−64 (*43*). This implies that the stability of CI2, and probably of the transition state for its formation, requires the protein to be intact, apart from the permissive 40−41 bond. Thus, for proteins such as CI2, there will be just a single nucleus, whereas for more structurally permissive proteins, such as the SH3 domain, the nucleus can change with the changing structure.

Nonnative Interactions in the Transition State of Unpermuted and Circular CI2 are Removed by Permutation. The main effect of the disulfide bond on the transition state of permuted CI2 is to make the structure more uniformly nativelike, i.e., change negative Φ_F-values to positive values. Side chains in several areas of unpermuted and circular CI2, such as Lys 2 and Arg 46, participate in local nonnative interactions in the transition state, but these interactions are not observed in permuted CI2. Nonnative interactions in CI2 may result from topological restraints, since they are exacerbated (their Φ_F-values become more negative) when the protein is circularized, but disappear completely upon subsequent permutation (Figure 3). The area around the termini and Arg 46 are "nonnative hot spots": circularization increases the nonnative character of Lys 2, as well as giving Pro 6, Val 60, and Tyr 42 negative Φ_F-values (Figure 3). Since several neighboring residues have negative Φ-values, the nonnative interactions are unlikely to be an artifact. Permuted CI2s increase in nativelike character around the termini and Arg 46 is offset by a modest decline in structure in other regions, such as the hydrophobic patch. Thus, there appears to be a conservation of total nativelike structure, as well as the folding nucleus, in the transition state of permuted CI2. The insensitivity of the folding pathway to changes in the reactive loop provides an opportunity for evolutionary optimization of the loop sequence (*43*).

For CI2, nonnative interactions are apparently unavoidable in the conformational search for a biologically active molecule. There is no evolutionary advantage in avoiding such interactions in the transition state, since the biologically active species is the native state rather than the transition state. Variants of CI2, in which these nonnative interactions are absent, are significantly worse inhibitors. Thus, permuted CI2 binds 25-fold less tightly than unpermuted CI2 to subtilisin BPN′ (D. E. Otzen, and A. R. Fersht, unpublished results), and the mutant RA46 of unpermuted CI2 binds 200-fold less tightly (A. G. Ladurner, D. E. Otzen, and A. R. Fersht, unpublished results).

Effective Concentration of Polypeptide Chains. The association, and concomitant folding, of fragments of CI2 is a second-order reaction whereas the folding of intact CI2 is first order. Comparison of the second-order and first-order rate constants for such processes allows the calculation of the loss of entropy on the two chains coalescing in the transition state for the formation of the complex (*5, 44, 45*). There is an assumption in such calculations that cleavage does not alter the inherent rate of the process, apart from the change in entropy. This assumption appears fully justified on comparing the association of peptides 1−40 and 41−64 of CI2, since the permutant folds with the same rate constant as wild-type (Table 4). The two peptides associate and fold with a second-order rate constant k_{inter} of 6×10^{-3} s^{-1} M^{-1}, while the intact protein folds with a first-order rate constant k_{intra} of 58 s^{-1}. This means that the effective molarity (k_{intra}/k_{inter}) is 10^4 M, i.e., the gain in entropy of activation on having the separate fragments linked is $Rln(10^4)$ or 18.3 cal mol^{-1} K^{-1}. Effective concentrations for pairs of Cys residues in proteins fall within the range of 10^2-10^5 M (*5*), but can be up to 10^{10} M for reactive groups in small molecules such as substituted benzene rings (*45*).

REFERENCES

1. Bowles, D. J., Marcus, S. E., Pappin, D., Findlay, J. B., Eliopoulos, E., Maycox, P. R., & Burgess, J. (1987) *J. Cell. Biol. 102*, 1284−1297.
2. Sheldon, P. S., Keen, J. N., & Bowles, D. J. (1996) *Biochem. J. 320*, 865−870.
3. Cunningham, B. A., Hemperly, J. J., Hopp, T. P., & Edelman, G. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3218−3222.
4. Goldenberg, D. P., & Creighton, T. E. (1983) *J. Mol. Biol. 165*, 407−413.
5. Goldenberg, D. P., & Creighton, T. E. (1984) *J. Mol. Biol. 179*, 527−545.
6. Luger, K., Hommel, U., Herold, M., Hofsteenge, J., & Kirschner, K. (1989) *Science 243*, 206−210.
7. Protasova, M. Y., Kireeva, M. L., Murzina, N. V., Murzin, A. G., Uversky, V. N., Gryznova, O. I., & Gudkov, A. T. (1994) *Protein Eng. 7*, 1373−1377.
8. Buchwalder, A., Szadkowski, H., & Kirschner, K. (1992) *Biochemistry 31*, 1621−1630.
9. Horlick, R. A., et al. (1992) *Protein Eng. 5*, 427−431.
10. Zhang, T., Bertelsen, E., Benvegnu, D., & Alber, T. (1993) *Biochemistry 32*, 12311−12318.
11. Yang, Y. R., & Schachmann, H. K. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11980−11984.
12. Graf, R., & Schachman, H. K. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11591−11596.
13. Hahn, D. P., Piotukh, K., Borriss, R., & Heinemann, U. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10417−10421.
14. Kreitman, R., Puri, R. K., & Pastan, I. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6889−6893.
15. Mullins, L. S., Wesseling, K., Kuo, J. M., Garrett, J. B., & Raushel, F. M. (1994) *J. Am. Chem. Soc. 116*, 5529−5533.

16. Vignais, M.-L., Corbier, C., Mulliert, G., Branlant, C., & Branlant, G. (1995) *Protein Sci. 4*, 994−1000.
17. Uversky, V. N., Kutyshenko, V. P., Protasova, N. Y., Rogov, V. V., Vassilenko, K. S., & Gudkov, A. T. (1996) *Protein Sci. 5*, 1844−1851.
18. Viguera, A. R., Blanco, F. J., & Serrano, L. (1995) *J. Mol. Biol. 247*, 670−681.
19. Viguera, A. R., Serrano, L., & Wilmanns, M. (1996) *Nat. Struct. Biol. 3*, 874−880.
20. Shakhnovich, E., Abkevich, V., & Ptitsyn, O. (1996) *Nature 379*, 96−98.
21. Shakhnovich, E. I. (1997) *Curr. Opin. Struct. Biol. 7*, 29−40.
22. Guo, Z., & Thirumalai, D. (1997) *Folding Des. 2*, 377−391.
23. McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261−269.
24. Harpaz, Y., elMasry, N. F., Fersht, A. R., & Henrick, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 311−315.
25. Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
26. Otzen, D. E., Itzhaki, L. S., ElMasry, N. F., Jackson, S. E., & Fersht, A. R. (1994) *Proc Natl. Acad. Sci. U.S.A. 91*, 10422−10425.
27. Itzhaki, L. S., Otzen, D. E., & Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
28. Li, A., & Daggett, V. (1995) *J. Mol. Biol. 257*, 412−427.
29. Neira, J. L., Davis, B., Ladurner, A. G., Buckle, A. M., de Prat Gay, G., & Fersht, A. R. (1996) *Folding Des. 1*, 189−208.
30. de Prat Gay, G., & Fersht, A. R. (1994) *Biochemistry 33*, 7964−7970.
31. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M., & Fersht, A. R. (1993) *Biochemistry 32*, 11259−11269.
32. Otzen, D. E., Rheinnecker, M., & Fersht, A. R. (1995) *Biochemistry 34*, 13051−13058.
33. Ellman, G. L. (1959) *Arch. Biochem. Biophys. 82*, 70−77.
34. Clarke, J., & Fersht, A. R. (1993) *Biochemistry 32*, 4322−4329.
35. Jackson, S. E., elMasry, N., & Fersht, A. R. (1993) *Biochemistry 32*, 11270−11278.
36. Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
37. Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261−287.
38. Hazes, B., & Dijkstra, B. W. (1988) *Protein Eng. 2*, 119−125.
39. Tanford, C. (1968) *Adv. Protein Chem. 23*, 121−217.
40. Fersht, A. R., Matouschek, A., & Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
41. Fersht, A. R., Itzhaki, L. S., ElMasry, N. F., Matthews, J. M., & Otzen, D. E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10426−10429.
42. van Mierlo, C. P. M., Kemmink, J., Neuhaus, D., Darby, N. J., & Creighton, T. E. (1994) *J. Mol. Biol. 235*, 1044−1061.
43. Ladurner, A. G., Itzhaki, L. S., de Prat Gay, G., & Fersht, A. R. (1997) *J. Mol. Biol. 273*, 317−329.
44. Page, M. I., & Jencks, W. P. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 1678−1683.
45. Kirby, A. J. (1980) *Adv. Phys. Org. Chem. 17*, 183−278.
46. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI980250G