Structure of the Hydrophobic Core in the Transition State for Folding of Chymotrypsin Inhibitor 2: A Critical Test of the Protein Engineering Method of Analysis†

Sophie E. Jackson,‡ Nadia elMasry, and Alan R. Fersht*

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

Received April 29, 1993; Revised Manuscript Received August 12, 1993®

ABSTRACT: Chymotrypsin inhibitor 2 (CI2) unfolds and refolds according to a simple two-state kinetic mechanism. The single rate-determining transition state may thus be studied by kinetics of both unfolding and refolding. This has allowed the direct testing of some facets of the protein engineering procedure (ϕ-value analysis). The structure of the hydrophobic core of CI2 in the transition state was analyzed from kinetic and thermodynamic measurements of guanidinium chloride-induced unfolding of 11 mutants and of their rates of refolding. In all cases, the strengths of the interactions measured from refolding kinetics in water are in excellent agreement with those measured from unfolding kinetics in guanidinium chloride solutions and extrapolated to zero molar denaturant. Changes in the free energies of unfolding on mutation, as well as other equilibrium properties calculated from the rate constants, are also in excellent agreement with those measured directly from equilibrium studies. These data provide further evidence for application of the principle of microscopic reversibility to aspects of protein folding in the presence of denaturant and the validity of extrapolation to the absence of denaturant. The edges of the hydrophobic core of CI2 are significantly weakened in the transition state, and, in many cases, the interactions are totally lost. The center of the core remains partially intact; the interaction energy is lowered by about 50%. In one case, Val → Ala38, a residue which lies on the edge of the core and whose side chain is 44% solvent-exposed, the interaction energy in the transition state (0.86 kcal mol^{-1}) is found to be larger than in the native state (0.46 kcal mol^{-1}), suggesting that this region is more buried in the transition state than in the native state. The results are consistent with the proposal that one of the final events in protein folding is the consolidation of the hydrophobic core and that the rate-determining step involves the close packing of the side chains within the hydrophobic core.

It is now widely accepted that the folding of proteins proceeds along a specific pathway or pathways rather than through a random search of all possible chain conformations (Levinthal, 1968). Many models have now been proposed for such pathways. Harrison and Durbin (1985) support the "jigsaw puzzle" model in which there exist a large number of parallel pathways; Kim and Baldwin (1982) support the "framework model" in which the first folding events are the formation of secondary structure elements; Chan and Dill (1990) support an alternative model in which hydrophobic collapse is an early event that is followed by secondary structure formation, and the diffusion–collision or subdomain model which assumes that the earliest folding events are the formation of structural subdomains with native conformation which then collide to form the final structure has been proposed by a number of groups (Karplus & Weaver, 1979; Oas & Kim, 1988; Stanley & Kim, 1990). In order to be able to distinguish between these various models, we need precise information on the interactions that are formed on the folding pathway, when such interactions are formed, and to what extent.

One approach has been to trap and characterize intermediates on the folding pathway. In some cases, this has given detailed information on the interactions formed along the folding pathway, for example, ribonuclease A (Udgaonkar & Baldwin, 1988), cytochrome *c* (Roder et al., 1988), and barnase (Bycroft et al., 1989). Another approach has been to use protein engineering methods and kinetics to analyze the interactions broken/formed in the transition state for folding (Matouschek et al., 1989, 1990, 1992). Kinetics is the only method for analyzing transition states and, in conjunction with protein engineering techniques, is a powerful method for determining accurate information on the energetics of interactions in the transition state. Each mutation acts as a reporter group for the events in its part of the structure. The strategy we have followed is as follows: (1) to find suitable simple interactions that may be disrupted by mutation without producing major structural changes; (2) to measure the changes in the free energy of unfolding of the mutants compared with wild-type (Jackson et al., 1993); (3) to measure unfolding and refolding kinetics of wild-type and mutant proteins to determine the fraction of the change in free energy of unfolding used in stabilizing the transition state (Matouschek et al., 1989; Fersht et al., 1992).

CI2 has been shown to follow, both kinetically and thermodynamically, the two-state model of protein folding; i.e., it has only one kinetically significant transition state (Jackson & Fersht, 1991a). This allows one to measure the strength of interactions in the transition state under denaturing conditions, from unfolding kinetics, and under native conditions, from refolding kinetics. Thus, we have independent means of verifying the results and a means of checking some of the assumptions made in the analysis of the results.

† S.E.J. is a William Stone Research Fellow, Peterhouse, Cambridge.
* Author to whom correspondence should be addressed.
‡ Present Address: Department of Chemistry, Harvard University, 12, Oxford St., Cambridge, MA 02138.
® Abstract published in *Advance ACS Abstracts,* October 1, 1993.

0006-2960/93/0432-11270$04.00/0

In this study, we analyze the kinetics of unfolding and refolding of 11 mutants in the hydrophobic core of CI2, and thereby characterize the energetics of hydrophobic interactions in the transition state for folding for this protein. The equilibrium analysis of these mutants has been discussed in the preceding paper (Jackson et al., 1993).

EXPERIMENTAL PROCEDURES

Materials

All chemicals are as described in the preceding paper (Jackson et al., 1993). Expression, purification, and description of the wild-type and mutant proteins are as described in the preceding paper (Jackson et al., 1993).

Equipment and General Procedures

[GdnHCl]-Jump Experiments. Reactions were followed with a Perkin-Elmer MPF-44B fluorescence spectrophotometer equipped with a rapid mixing head. The mixing device contained a T-jet mixing chamber followed by a 30-ms delay loop, ensuring complete mixing of the solutions before observation. A Hellma flow-through compact cell (10 mm × 3 mm × 3 mm) was used. The solutions were driven through the mixing chamber manually from two Hamilton syringes, resulting in a mixing ratio of 1:10. The observation cell and reservoir syringes were thermostated separately using two water baths, and the temperature of each was monitored using an Edale Instrument thermometer. Temperatures were kept within ±0.1 °C. Data were acquired with a Tandon Target microcomputer, a DT2801 data translation board, and the Bio-kine data acquisition software package by Bio-logic, and analyzed using the program *Enzfitter* (Elsevier Biosoft, Cambridge) by R. J. Leatherbarrow.

pH-Jump Experiments. An Applied Photophysics stopped-flow spectrophotometer, Model SF 17MV, was used to monitor the reaction. A 1:1 mixing ratio was used in all pH-jump experiments. The temperature of the cell and reservoir synringes was maintained by thermostating with a Grant LTD6 water bath. Temperatures were maintained to ±0.1 °C on the basis of an internal temperature probe in the stopped-flow apparatus which had previously been calibrated against an Edale Instrument thermometer. A Radiometer PHM64 Research pH meter was used for pH measurements.

Unfolding Kinetics. Unfolding kinetics were measured by [GdnHCl]-jump experiments as described above. Unfolding was initiated by diluting 1 volume of the aqueous protein solution (approximately 45 μM, 50 mM MES, pH 6.3) into 10 volumes of concentrated guanidinium chloride solution (containing 50 mM MES, pH 6.3); concentrations were such that the final concentrations were between 4.0 and 7.0 M.

Refolding Kinetics. Refolding kinetics were measured using both [GdnHCl]-jump and pH-jump experiments as described above.

(i) [GdnHCl]-Jump. Refolding was initiated, after pre-equilibration of the protein in 6.5 M GdnHCl/50 mM MES, pH 6.3, by diluting it, 11-fold, into low concentrations of GdnHCl/50 mM MES, pH 6.3. Final guanidinium chloride concentrations were between 0.59 and 3.5 M.

(ii) pH-Jump. Points between 0 and 0.6 M GdnHCl were obtained by a pH-jump experiment. Protein was initially acid-denatured by lowering the pH of the solution to 1.7 by addition of 5 M HCl. At this pH, it was found that the protein is completely denatured as determined by fluorescence spectroscopy. The subsequent addition of GdnHCl makes no further difference to the fluorescence spectrum. The protein was refolded by rapid mixing (1:1) with a strongly buffered solution at high pH. The final pH of the solution was pH 6.3, 50 mM MES. An Applied Photophyics stopped-flow Model SF 17MV spectrophotometer was used to monitor the fast reaction. The pH-jump experiment was performed in the absence and in the presence of low concentrations of denaturant. All kinetic experiments were performed at 25 °C.

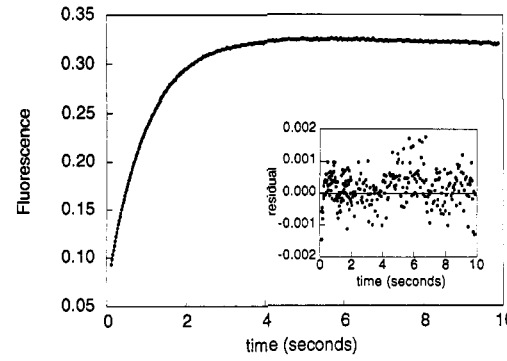

FIGURE 1: Unfolding kinetic trace of wild-type CI2. Unfolding of wild-type CI2 (approximately 45 μM, 50 mM MES, pH 6.3) was initiated by diluting 1 volume of protein into 10 volumes of concentrated guanidinium chloride solution (containing 50 mM MES, pH 6.3); concentrations were such that the final concentration was 7.0 M. The unfolding was followed by monitoring the change in the intrinsic fluorescence of CI2 (excitation wavelength was 280 nm, emission wavelength was 356 nm). Insert: residuals after data were fitted to a single-exponential process with linear drift and offset (eq 1).

RESULTS

Analysis of Unfolding Kinetics. Unfolding of CI2 is accompanied by a large increase in fluorescence at 356 nm. A typical unfolding trace is shown in Figure 1. The fluorescence traces were analyzed using the nonlinear regression analysis program *Enzfitter*. The data were fitted to an equation describing a single-exponential process with linear drift and offset (eq 1) where $F(t)$ is the fluorescence at time

$$F(t) = A_0[1 - \exp(-k_u t)] - mt + C \quad (1)$$

t, A_0 is the amplitude, k_u is the rate constant, m is the slope of the drift, and C is an offset. The drift, when present, is very small and results from base-line instability. Plots of the natural logarithm of the rate constants of unfolding against the final guanidinium chloride concentration are linear (Figure 2), conforming to eq 2 where k_u is the rate constant of unfolding

$$\ln k_u = \ln k_u^{H_2O} + m_{k_u}[\text{GdnHCl}] \quad (2)$$

at a given GdnHCl concentration, $k_u^{H_2O}$ is the rate constant of unfolding in water, m_{k_u} is the slope, and [GdnHCl] is the final guanidinium chloride concentration. The natural logarithms of the rate constants in water, $\ln k_u^{H_2O}$, and the slopes, m_{k_u}, for all the mutants are tabulated in Table I. The standard errors (all errors are calculated from the best fit of the data and are not standard deviations from repetitive runs, unless otherwise stated) are also given and are typically ±0.12, on average for $\ln k_u^{H_2O}$, and ±0.02 mol^{-1}, on average for m_{k_u}. If the plots are extrapolated to 4 M GdnHCl instead of water (0 M GdnHCl), then the errors in $\ln k_u^{4M}$ are significantly lower, typically ±0.03, because of the much shorter extrapolation. The rates of unfolding of the mutants are faster than those for wild-type (Figure 2), as expected for destabilizing mutations. The effect of mutation on the energy of the transition state of unfolding can be calculated using transition-

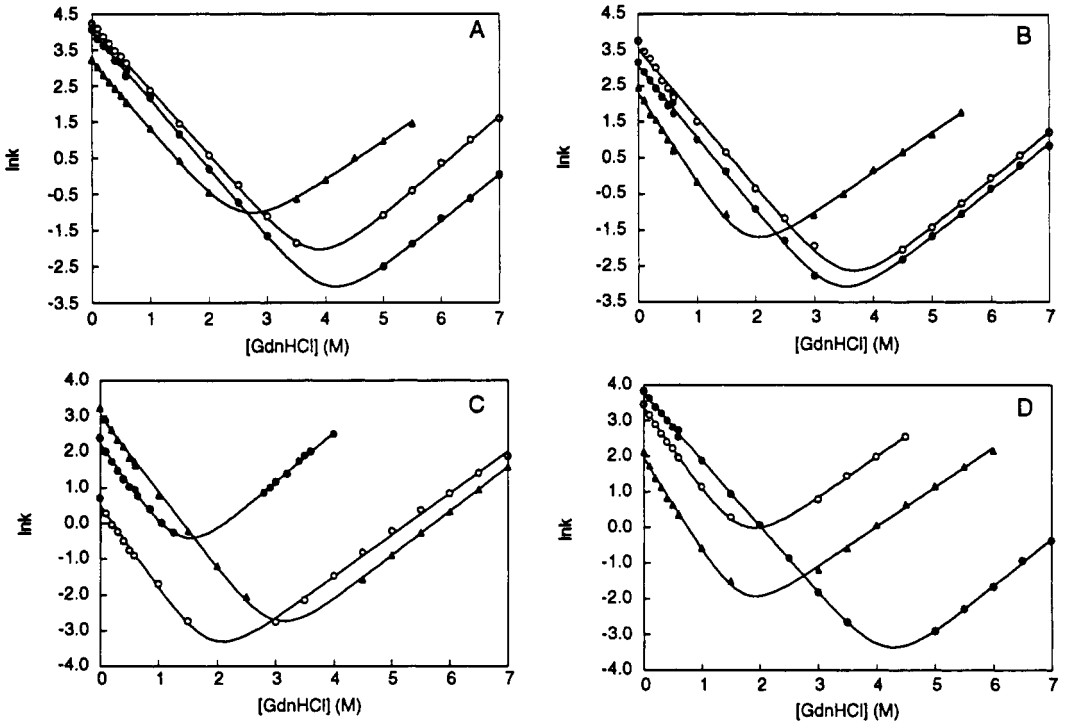

FIGURE 2: Combined unfolding and refolding kinetics. Plots of ln k, where k is the rate constant for unfolding or refolding, *versus* guanidinium chloride concentration. The solid lines show the best fit of the data to the two-state model (eq 10). (A) (Filled circles) Wild-type; (filled triangles) LA27; (open circles) VA38. (B) (Filled circles) IV39; (open circles) IV48; (filled triangles) IA48. (C) (Filled circles) VA66; (open circles) LA68; (filled triangles) VA70. (D) (Filled circles) IV76; (filled triangles); IA76 (open circles) IA48IV76.

Table I: Rate and Thermodynamic Data from Unfolding Kinetics of Hydrophobic Core Mutants

mutant	m_{k_u} (kcal mol^{-2})	ln $k_u^{H_2O}$	$\Delta\Delta G_{\ddagger-F}^{H_2O}$ (kcal mol^{-1})[a]	$\Delta\Delta G_{\ddagger-F}^{4M}$ (kcal mol^{-1})[b]	$\Delta\Delta G_{U-F}^{4M}$ (kcal mol^{-1})[b]	$\Delta\Delta G_{U-F}^{[G]_{50\%}}$ (kcal mol^{-1})[b]	$\phi_u^{H_2O}$ (water)[c]	ϕ_u^{4M} (4 M GdnHCl)[d]
wild-type	1.30 ± 0.01	−9.04 ± 0.07						
LA27	1.05 ± 0.03	−4.42 ± 0.14	2.73 ± 0.09	2.24 ± 0.02	2.95 ± 0.19	2.64 ± 0.06	1.03 ± 0.04	0.76 ± 0.05
VA38[e]	1.35 ± 0.02	−7.80 ± 0.13	0.73 ± 0.09	0.86 ± 0.03	0.48 ± 0.08	0.46 ± 0.07	1.59 ± 0.31	1.78 ± 0.30
IV39	1.27 ± 0.02	−8.02 ± 0.11	0.60 ± 0.08	0.57 ± 0.02	1.31 ± 0.10	1.27 ± 0.06	0.47 ± 0.06	0.44 ± 0.04
IV48	1.32 ± 0.01	−8.00 ± 0.09	0.61 ± 0.07	0.73 ± 0.03	1.12 ± 0.12	1.09 ± 0.07	0.56 ± 0.07	0.65 ± 0.07
IV48	1.12 ± 0.03	−4.41 ± 0.12	2.74 ± 0.08	2.34 ± 0.02	4.12 ± 0.39	3.84 ± 0.09	0.71 ± 0.03	0.57 ± 0.05
VA66	1.41 ± 0.02	−3.08 ± 0.08	3.53 ± 0.06	3.81 ± 0.02	4.45 ± 0.59	4.88 ± 0.21	0.72 ± 0.03	0.86 ± 0.11
LA68	1.14 ± 0.03	−5.98 ± 0.16	1.81 ± 0.10	1.40 ± 0.01	3.92 ± 0.25	3.82 ± 0.09	0.47 ± 0.03	0.36 ± 0.02
VA70	1.25 ± 0.01	−7.20 ± 0.05	1.09 ± 0.05	1.00 ± 0.02	2.17 ± 0.21	1.95 ± 0.07	0.56 ± 0.03	0.46 ± 0.04
IV76[f]	1.28 ± 0.03	−9.31 ± 0.15	−0.16 ± 0.10	−0.20 ± 0.03	−0.20 ± 0.10	−0.21 ± 0.10		
IA76	1.17 ± 0.04	−2.67 ± 0.16	3.77 ± 0.10	3.46 ± 0.02	4.25 ± 0.48	4.25 ± 0.12	0.89 ± 0.03	0.82 ± 0.09
IA48/IV76	1.14 ± 0.03	−4.51 ± 0.14	2.68 ± 0.09	2.29 ± 0.02	3.82 ± 0.31	4.05 ± 0.10	0.66 ± 0.03	0.60 ± 0.05

[a] Values are calculated using eq 3 at 0 and 4 M GdnHCl, respectively. [b] Values are calculated from equilibrium experiments, see Jackson et al. (1993)]. [c] Calculated from eq 4 using $\Delta\Delta G_{\ddagger-F}^{H_2O}$ and $\Delta\Delta G_{\ddagger-F}^{[G]_{50\%}}$. [d] Calculated from eq 4 using $\Delta\Delta G_{\ddagger-F}^{4M}$ and $\Delta\Delta G_{U-F}^{4M}$. [e] ϕ_u is greater than 1 for this for this mutant as $\Delta\Delta G_{\ddagger-F} > \Delta\Delta G_{\ddagger-F}$. [f] ϕ_u not calculated for this mutant as $\Delta\Delta G_{\ddagger-F}$ and $\Delta\Delta G_{U-F}$ are close to 0.

state theory. In order to obtain information about the structure of the transition state, the stability of the transition state of the mutant protein relative to that of the wild-type protein is calculated from eq 3 where $\Delta\Delta G_{\ddagger-F}$ is the difference in energy

$$\Delta\Delta G_{\ddagger-F} = -RT \ln(k_u/k_u') \quad (3)$$

of the transition state of unfolding relative to the folded state between wild-type and mutant and k_u' is the rate constant of unfolding for mutant protein. The values for $\Delta\Delta G_{\ddagger-F}$, calculated at both 0 M and 4 M GdnHCl, for hydrophobic core mutants are given in Table I. The standard error of the energies calculated in water, $\Delta\Delta G_{\ddagger-F}^{H_2O}$, is on average ±0.08 kcal mol^{-1}, while the energies calculated at 4 M, $\Delta\Delta G_{\ddagger-F}^{4M}$, are more precise at ±0.02 kcal mol^{-1}. The values of $\Delta\Delta G_{\ddagger-F}$ at 4 M GdnHCl are close to or differ only slightly from those calculated from the rate constants extrapolated to 0 M GdnHCl. In order to analyze the structure of the transition state of unfolding, a ratio ϕ_u is introduced:

$$\phi_u = \Delta\Delta G_{\ddagger-F}/\Delta\Delta G_{U-F} \quad (4)$$

where $\Delta\Delta G_{U-F}$ is the difference in the free energy of unfolding between wild-type and mutant. This value can be calculated from GdnHCl-induced denaturation experiments performed under equilibrium conditions, as discussed in the preceding paper (Jackson et al., 1993). This value may be calculated in water, $\Delta\Delta G_{U-F}^{H_2O}$, at 4 M GdnHCl, $\Delta\Delta G_{U-F}^{4M}$, or at the midpoint between the midpoints of unfolding of wild-type and mutant protein, $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$. The relative merit of the three methods of calculating $\Delta\Delta G_{U-F}$ and the errors of the resulting values are discussed in great detail in the preceding paper (Jackson et al., 1993). We have calculated ϕ_u for two sets of conditions: first, in water, using $\Delta\Delta G_{\ddagger-F}^{H_2O}$. The value for $\Delta\Delta G_{U-F}$ used is $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$, which was shown to be in excellent agreement with values measured directly in water by calorimetry, but which has lower errors (Jackson et al., 1993). This gives $\phi_u^{H_2O}$ shown in Table I. The second

condition is in 4 M GdnHCl using $\Delta\Delta G_{\ddagger-F}{}^{4M}$ and $\Delta\Delta G_{U-F}{}^{4M}$. This gives the value $\phi_u{}^{4M}$ also shown in Table I. The differences in $\phi_u{}^{H_2O}$ and $\phi_u{}^{4M}$ result from the small differences in m_{k_u} and, in most cases, are within experimental error. The values of ϕ_u at 4 M GdnHCl are directly relevant to unfolding since these are calculated from values of $\Delta\Delta G_{\ddagger-F}$ that are measured under the same conditions as those for $\Delta\Delta G_{U-F}$ [see Jackson et al. (1993)]. The values in water may be directly compared with data for refolding in water.

Solvent Accessibility of the Transition State for Unfolding. The slopes, m_{k_u}, are close to the wild-type slope of 1.30 ± 0.01 (Table I). This indicates that both the nature of the transition state and the nature of the folded state are not altered drastically by the mutations and that the unfolding pathway remains essentially unaltered. The fractional increase in exposure of a protein in the transition state is given by $m_u^{\ddagger}/m$ (Tanford, 1970). Using the wild-type slope (m_{k_u} = 1.30 ± 0.01), which gives a value of 0.77 kcal mol^{-1} for $m_u^{\ddagger}$, and the average value of the slope m (±standard error) = 1.93 ± 0.025, from equilibrium unfolding studies (Jackson et al., 1993) this gives a fractional increase in solvent exposure of the wild-type protein on going to the transition state for unfolding of 0.40 ± 0.01.

Analysis of Refolding Kinetics. The refolding of chymotrypsin inhibitor 2 (CI2) is, at least, a triphasic process. The rate constants are 53 s^{-1} for the major phase (77% of the total amplitude) and 0.43 and 0.024 s^{-1} for the slower phases (23% of the total amplitude) at 25 °C and pH 6.3. The multiphasic nature of the refolding reaction results from heterogeneity in the denatured state because of proline isomerization. The fast phase corresponds to the refolding of the fraction of protein that has all its prolines in a native *trans* conformation in the denatured state. It is not catalyzed by peptidylprolyl isomerase. The rate-limiting step of folding for the slower phases, however, is proline isomerization, and they are both catalyzed by peptidylprolyl isomerase (Jackson & Fersht, 1991b). The analysis of the energetics and pathway in this study concerns only the major fast refolding phase which represents the folding of the *all-trans*-proline unfolded state.

Refolding of CI2 is accompanied by a large decrease in the fluorescence at 356 nm. The rate constant for the fast phase is more than 2000-fold faster than that for the slowest phase. Over the time course of the experiment, typically 0–500 ms, the slowest proline-dependent phase makes no contribution to the overall change in fluorescence. Under such conditions, the data can be fitted to an equation describing a double-exponential process with offset (eq 5, see Figure 3) where $F(t)$

$$F(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t) + C \qquad (5)$$

is the fluorescence at time t, A_1 and A_2 are the amplitudes of the different phases, k_1 and k_2 are the rate constants, and C is an offset. k_f, the refolding rate constant, is taken as the faster of the two rate constants; the second rate constant is that of a proline-dependent phase (Jackson & Fersht, 1991b). Data are treated the same whether measured by [GdnHCl]-jump or by pH-jump experiment. Plots of the natural logarithm of the rate constants of refolding against the final guanidinium chloride concentration, obtained by [GdnHCl]-jump experiment or by pH-jump experiment, are linear (Figure 2) following eq 6 where k_f is the rate constant of folding at

$$\ln k_f = \ln k_f{}^{H_2O} + m_{k_f}[\text{GdnHCl}] \qquad (6)$$

a given GdnHCl concentration, $k_f{}^{H_2O}$ is the rate constant of folding in water, m_{k_f} is the slope, and [GdnHCl] is the final

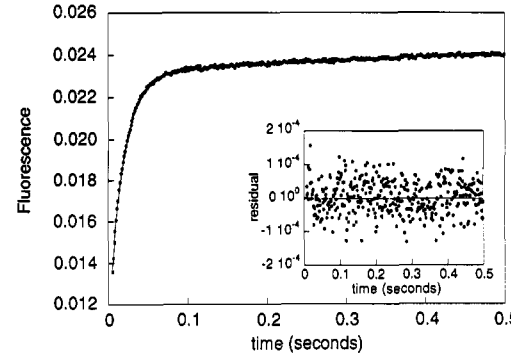

FIGURE 3: Refolding kinetic trace of wild-type CI2. Refolding was initiated by rapidly mixing a sample of denatured protein (pH 1.7) (1:1) with a strongly buffered solution at high pH such that the final pH of the solution was 6.3, 50 mM MES. The y axis is inverted by the stopped-flow apparatus; therefore, an increase in signal as shown in the figure represents a decrease in the actual fluorescence. Insert: residuals after data were fitted to a double-exponential process with offset (eq 5).

guanidinium chloride concentration. The natural logarithms of the rate constants in water, ln $k_f{}^{H_2O}$, and the slopes, m_{k_f}, measured both by [GdnHCl]-jump and by pH-jump, for all the mutants are tabulated in Table II. In addition, the values obtained by combining both sets of data are also given. The standard errors are also given and are typically ±0.06 on average for ln $k_f{}^{H_2O}$ and ±0.05 on average for m_{k_f} for [GdnHCl]-jump experiments, and ±0.04 for ln $k_f{}^{H_2O}$ and ±0.05 for m_{k_f} when data from [GdnHCl]-jump and pH-jump experiments are combined.

The kinetics of refolding for each mutant were measured by two procedures: (i) refolding after GdnHCl denaturation by diluting out the denaturant; (ii) refolding after acid denaturation by restoring to high pH. This allowed monitoring the kinetics on increasing the concentration of GdnHCl from 0 M. For several of the mutants, the two sets of data merge perfectly. For some, there is a slight deviation, which may be real (Figure 2). In these cases, however, the deviations are small in absolute terms, as can be seen by the values of ln $k_f{}^{H_2O}$ and m_{k_f} calculated from just the experiments in GdnHCl and those from the combined data (Table II).

The effect of mutation on the energy of the transition state of folding can be calculated using transition-state theory in a similar manner to the analysis of the unfolding data. In order to obtain information about the structure of the transition state, the stability of the transition state of mutant protein relative to that of wild-type protein is calculated from eq 7

$$\Delta\Delta G_{\ddagger-U} = -RT \ln(k_f/k_f') \qquad (7)$$

where $\Delta\Delta G_{\ddagger-U}$ is the difference in energy of the transition state of folding relative to the unfolded state between wild-type and mutant and k_f' is the rate constant of folding for mutant protein. The values for $\Delta\Delta G_{\ddagger-U}$ for hydrophobic core mutants are given in Table II. The standard error of the value for $\Delta\Delta G_{\ddagger-U}{}^{H_2O}$ calculated from the [GdnHCl]-jump experiment is ±0.08 kcal mol^{-1}, and from combining [GdnHCl]-jump and pH-jump data, it is ±0.05 kcal mol^{-1}. In order to analyze the structure of the transition state for folding, a ratio ϕ_f is defined as

$$\phi_f = \Delta\Delta G_{\ddagger-U}/\Delta\Delta G_{U-F} \qquad (8)$$

$\Delta\Delta G_{U-F}{}^{[G]_{50\%}}$ is used as a measure of the difference in the free energy of unfolding between wild-type and mutant. Although measured at a [GdnHCl] which is midway between the midpoints of unfolding for wild-type and mutant, it has been

Table II: Rate and Thermodynamic Data from Refolding Kinetics of Hydrophobic Core Mutants

	[GdnHCl]-jump data				combined [GdnHCl]- and pH-jump data				equilibrium
mutant	m_{k_f} (mol^{-1})	ln $k_f^{H_2O}$	$\Delta\Delta G_{\ddagger-U}^{H_2O}$ (kcal mol^{-1})[a]	ϕ_f (water)[b]	m_{k_f} (kcal mol^{-2})	ln $k_f^{H_2O}$	$\Delta\Delta G_{\ddagger-U}^{H_2O}$ (kcal mol^{-1})[a]	ϕ_f (water)[b]	$\Delta\Delta G_{U-F}^{[G]_{50\%}}$ (kcal mol^{-1})[c]
wild-type	1.87 ± 0.03	3.95 ± 0.05			1.90 ± 0.02	4.02 ± 0.02			
LA27	1.77 ± 0.01	3.09 ± 0.01	0.51 ± 0.03	0.19 ± 0.01	1.83 ± 0.02	3.17 ± 0.02	0.50 ± 0.02	0.19 ± 0.01	2.64 ± 0.06
VA[38][d]	1.73 ± 0.02	4.07 ± 0.04	−0.07 ± 0.04	−0.15 ± 0.09	1.79 ± 0.01	4.21 ± 0.02	−0.11 ± 0.02	−0.24 ± 0.05	0.46 ± 0.07
IV39	1.89 ± 0.02	2.91 ± 0.05	0.62 ± 0.04	0.48 ± 0.04	1.94 ± 0.02	3.01 ± 0.03	0.60 ± 0.02	0.47 ± 0.03	1.27 ± 0.06
IV48	1.76 ± 0.03	3.26 ± 0.06	0.41 ± 0.05	0.37 ± 0.05	1.88 ± 0.05	3.49 ± 0.06	0.31 ± 0.04	0.29 ± 0.04	1.09 ± 0.07
IV48	1.92 ± 0.07	1.80 ± 0.08	1.27 ± 0.06	0.33 ± 0.02	2.32 ± 0.10	2.24 ± 0.07	1.05 ± 0.04	0.27 ± 0.01	3.84 ± 0.09
VA66	1.71 ± 0.07	1.84 ± 0.07	1.25 ± 0.05	0.26 ± 0.02	2.06 ± 0.09	2.16 ± 0.06	1.10 ± 0.04	0.23 ± 0.01	4.88 ± 0.21
LA68	2.01 ± 0.04	0.28 ± 0.04	2.17 ± 0.04	0.57 ± 0.02	2.20 ± 0.09	0.46 ± 0.06	2.11 ± 0.04	0.55 ± 0.02	3.82 ± 0.09
VA70	1.98 ± 0.06	2.82 ± 0.10	0.67 ± 0.07	0.34 ± 0.04	2.10 ± 0.03	3.01 ± 0.05	0.60 ± 0.03	0.31 ± 0.02	1.95 ± 0.07
IV76[e]	1.80 ± 0.01	3.65 ± 0.03	0.18 ± 0.03		1.85 ± 0.01	3.75 ± 0.02	0.16 ± 0.02		−0.21 ± 0.10
IA76	1.85 ± 0.09	3.04 ± 0.10	0.54 ± 0.07	0.13 ± 0.02	2.10 ± 0.07	3.30 ± 0.05	0.43 ± 0.03	0.10 ± 0.01	4.25 ± 0.12
IA48/IV76	2.02 ± 0.12	1.50 ± 0.13	1.45 ± 0.08	0.36 ± 0.02	2.39 ± 0.10	1.89 ± 0.07	1.26 ± 0.04	0.31 ± 0.01	4.05 ± 0.10

[a] Calculated using eq 7. [b] Calculated using $\Delta\Delta G_{\ddagger-U}^{H_2O}$ and $\Delta\Delta G_{U-F}^{[G]_{50\%}}$ (eq 8). [c] Data taken from preceding paper (Jackson et al., 1993). [d] ϕ_f is negative for this mutant as $\Delta\Delta G_{\ddagger-F} > \Delta\Delta G_{U-F}$. [e] ϕ_f not calculated for this mutant as $\Delta\Delta G_{\ddagger-F}$ and $\Delta\Delta G_{U-F}$ are close to 0.

shown, from a comparison to calorimetric results obtained in aqueous solution, that $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ is also a very good measure of the difference in the free energy of unfolding between wild-type and mutant in water (Jackson et al., 1993). In addition, $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ can be measured with a much higher degree of accuracy than $\Delta\Delta G_{U-F}^{H_2O}$ which has large errors because of the extrapolation from 3–4 M GdnHCl to water. The values of ϕ_f are shown in Table II.

CI2 has been shown, both thermodynamically and kinetically, to follow the two-state model of protein folding (Jackson & Fersht, 1991a). Therefore, $\Delta\Delta G_{\ddagger-U}$ is related to $\Delta\Delta G_{\ddagger-F}$ by $\Delta\Delta G_{\ddagger-U} = \Delta\Delta G_{U-F} - \Delta\Delta G_{\ddagger-F}$, and so:

$$\phi_u = 1 - \phi_f \quad (9)$$

Unfolding and refolding experiments have previously been performed at the same final guanidinium chloride concentration within the transition region, and the rate constants have been shown to depend only on the final conditions; thus, the folding of CI2 has been shown to be reversible (Jackson & Fersht, 1991a,b).

Analysis of Two-State Behavior. The complete kinetics of folding and unfolding can be fitted to eq 10, which is derived from eq 2 and 6, and based on a two-state transition where

$$\ln k = \ln[k_f^{H_2O} \exp(-m_{k_f}[\mathrm{GdnHCl}]) + k_u^{H_2O} \exp(m_{k_u}[\mathrm{GdnHCl}])] \quad (10)$$

k is the rate constant of unfolding or refolding at a particular GdnHCl concentration. For wild-type and many of the hydrophobic core mutants, the kinetic data for unfolding and refolding fit to this two-state model. The calculated fit of the data to eq 10 is shown by the solid curves in Figure 2. The values of $k_f^{H_2O}$, $k_u^{H_2O}$, m_{k_f}, and m_{k_u}, calculated from the kinetic data using eq 10, can be used to calculate the equilibrium values of $\Delta G_{U-F}^{H_2O}$, $\Delta\Delta G_{U-F}^{H_2O}$, $[\mathrm{GdnHCl}]_{50\%}$ (the concentration of guanidinium chloride at which 50% of the protein is denatured), and m. However, the equilibrium due to proline isomerization has to be taken into account:

$$U_c \underset{23\%}{\overset{K_{iso}}{\rightleftharpoons}} U_t \underset{77\%}{\overset{k_f^{H_2O}}{\rightleftharpoons}} N \quad (k_u^{H_2O})$$

where U_t is the denatured protein with all its prolines in a *trans* conformation, U_c is the denatured protein with a proline in a *cis* conformation, and N is the native state, which has all its prolines in a *trans* conformation. Defining the equilibrium constant for isomerization, K_{iso}, as $[U_c]/[U_t]$ and the apparent equilibrium constant for unfolding calculated from the kinetic experiments, K, as $k_f^{H_2O}/k_u^{H_2O}$, it follows that the equilibrium constant for unfolding, K_U, is $K/(1 + K_{iso})$. From the relative amplitudes of the slow and fast phases, 23% and 77%, respectively, $K_{iso} = 0.299$. Similarly, m can be calculated from the kinetic data using the relationship $m = 0.592(m_{k_f} + m_{k_u})$. $[\mathrm{GdnHCl}]_{50\%}$ can be calculated according to eq 11:

$$[\mathrm{GdnHCl}]_{50\%} = \frac{\ln(k_f^{H_2O}/k_u^{H_2O})}{(m_{k_f} + m_{k_u})(1 + K_{iso})} \quad (11)$$

Kinetic Two-State Behavior of Hydrophobic Core Mutants. Figure 2 shows the unfolding and refolding, both [GdnHCl]-jump and pH-jump, kinetics for all mutants fitted to the two-state model (eq 10). The best fit is shown by the solid line. It can be seen that wild-type, Leu → Ala27, Val → Ala38, Ile → Ala39, and Ile → Val76 fit the two-state model extremely well. The other mutants do not fit the model so well, particularly at low [GdnHCl] where the measured points are slightly higher than the calculated fit. A comparison of the values for $\Delta G_{U-F}^{H_2O}$ and $\Delta\Delta G_{U-F}^{H_2O}$, calculated from the two-state fit of the kinetic data, and the values measured directly from equilibrium experiments (Jackson et al., 1993) shows that, in almost all cases, the values are within error. The correlation between kinetic and equilibrium data is shown in Figure 4A,B. The slopes of the plots are 0.94 and 0.97, and the correlation coefficients are 0.97 and 0.99, for $\Delta G_{U-F}^{H_2O}$ and $\Delta\Delta G_{U-F}^{H_2O}$, respectively. The values for Leu → Ala27 do not fall within experimental error but are, nevertheless, close. Comparison of the m values calculated from kinetic and equilibrium data indicates that, in most cases, the values are the same within error. The value for Leu → Ala27 is again not quite within the error, and for Ile → Ala48, Val → Ala66, and Ile → Ala48/Ile → Val76, the m values calculated from kinetic data are higher than expected when the pH-jump data are included.

The best correlation (see Figure 4C) is found between the $[\mathrm{GdnHCl}]_{50\%}$ calculated from the kinetic data and the equilibrium data. Most of the values lie well within the experimental error, only the very destabilized double mutant Ile → Ala48/Ile → Val76 is slightly out. The slope for the plot is 1.09, and the correlation coefficient is 1.00.

Thus, we have shown that all the hydrophobic core mutants still follow the two-state model of protein folding, both thermodynamically [see Jackson et al. (1993)] and kinetically. This indicates that none of the mutations have a significant effect on the pathway of folding of CI2.

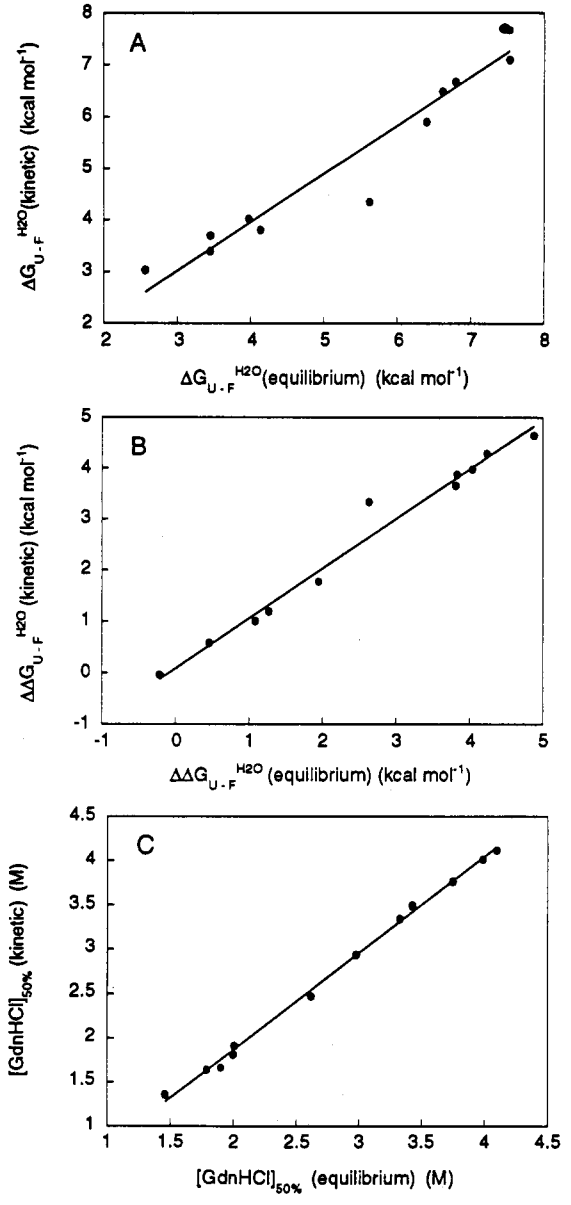

FIGURE 4: Comparison of the results obtained from equilibrium experiments (Jackson et al., 1993) with those calculated from kinetic experiments and based on the best fit of the data to eq 10. (A) Comparison of $\Delta G_{U-F}^{H_2O}$ calculated from equilibrium (abscissa) and kinetic (ordinate) experiments. (B) Comparison of $\Delta\Delta G_{U-F}^{H_2O}$ calculated from equilibrium (abscissa) and kinetic (ordinate) experiments. (C) Comparison of $[GdnHCl]_{50\%}$ calculated from equilibrium (abscissa) and kinetic (ordinate) experiments.

Interpretation of ϕ Values. We discuss the ϕ values in terms of ϕ_f, either measured directly or taken from eq 9 ($1 - \phi_u$). There are two extreme values of ϕ_f that may be interpreted in a simple manner. A value of $\phi_f = 1$ implies that the structure at the site of mutation is as folded in the transition state as it is in the folded state. Conversely, $\phi_f = 0$ shows that the structure at the site of mutation is as unfolded in the transition state as it is in the unfolded structure. In general, fractional values of ϕ_f are more difficult to interpret. However, in the case of mutations such as those made here, i.e., nondisruptive deletion of side chains without access of water to the site of mutation (hydrophobic core residues), then ϕ_f is a measure of the gain of van der Waals interaction energies (Matouschek et al., 1989; Fersht et al., 1992). This assumes the difference in the solvation energy of the wild-type and mutant side chain is similar. In the case of a mutation of a hydrophobic residue to another hydrophobic residue, this assumption is valid. The theory, assumptions, and limitations of this approach have been discussed elsewhere (Fersht et al., 1992).

ϕ Value Analysis of Mutations. The values of ϕ_f calculated from unfolding and refolding data are summarized in Table III, as are the locations of the mutated residue and the major interactions the side chain makes. The values of ϕ_f for the hydrophobic core mutants fall into two categories: those that have values close to 0 and those that have fractional values. As yet, we have not found any mutations which have values close to 1. In addition to the values calculated for the actual mutations, a fine-structure analysis has been performed for the mutations in which one side chain has been subjected to a variety of substitutions. For example, when a large hydrophobic side chain is deleted, e.g., Ile → Ala, the resulting ϕ_f is the average of that of all the interactions that are made by $C^{\gamma 1}$, $C^{\gamma 2}$, and $C^{\delta 1}$. In this case, we can attribute ϕ_f values to different parts of the side chain by analyzing first the mutation Ile → Val, in which only the $C^{\delta 1}$ is removed, and then the mutation Ile → Ala, in which $C^{\gamma 1}$, $C^{\gamma 2}$, and $C^{\delta 1}$ methyl(ene) groups are removed. We can combine the data to give the change for Val → Ala in which only $C^{\gamma 1}$ and $C^{\gamma 2}$ are removed. In this way, individual ϕ values may be attributed to different parts of a single side chain in the transition state (Table IV). The structural features of the residues that have been mutated are discussed fully in the preceding paper (Jackson et al., 1993).

(i) $\phi_f = 0$; Interactions That Are Not Made in the Transition State. *(a) Leu → Ala27.* Leu27 is located between a type III reverse turn (Trp24–Leu27) and a type II reverse turn (Leu27–Lys30) just before the α-helix. It packs against the side chain of Trp24 (β-strand$_1$), Lys30 (also in the turn), several side chains in the α-helix (Ala35, Val38, and Ile39), and Pro80 (β-strand$_4$). The side chain of Leu27 is almost completely buried; it has a solvent-accessible surface area of only 1.4 Å^2, but it lies on the edge of the hydrophobic core. The mutation Leu → Ala27 gives information about the integrity of the reverse turn, and its packing against the α-helix and β-strands 1 and 4. ϕ_f is close to zero, indicating that the interactions that this side chain makes in the native state are not made in the transition state. These results suggest that the edge of the core in the region of Leu27 is not formed in the transition state.

(b) Val → Ala66. Val66 is the first residue in β-strand$_3$ and, therefore, lies on the edge of the β-sheet. The side chain of Val66 interacts with Trp24 (β-strand$_1$), with Ile39 and Lys43 (α-helix), and to a lesser extent with Pro80 and Val82 (β-strand$_4$) and Ala46 (turn$_3$). Thus, the mutation Val → Ala66 monitors the interaction between β-strand$_3$ and (a) the α-helix and (b) β-strand$_1$. ϕ_f is close to zero, indicating that the interactions this side chain makes in the native state are not formed in the transition state. This suggests that this region of the core is not highly structured in the transition state.

(c) Ile → Val76; Ile → Ala76. Ile76 is the second residue of β-strand$_4$, and lies on the edge of the β-sheet and the hydrophobic core. The $C^{\delta 1}$ methyl group interacts with Leu68 and Val70 (β-strand$_3$) and Val32 and Ala35 (α-helix). The mutation Ile → Val76 therefore probes the interaction of β-strand$_4$ with β-strand$_3$ and the α-helix. However, the ϕ_f values are difficult to measure accurately for this mutant because $\Delta\Delta G_{\ddagger-F}$ and $\Delta\Delta G_{U-F}$ are both close to zero. The $C^{\gamma 1}$ and $C^{\gamma 2}$ methyl groups of Ile76 interact with the side chains of Pro80 (β-strand$_4$), Leu68 (β-strand$_3$), and Ala35 (α-helix). Therefore, the mutation Ile → Ala76 probes the interaction

Table III: Summary of Locations and Interactions of the Mutated Residues and Their ϕ_u and ϕ_f Values

mutation	location	interactions[a]	$(1-\phi_u) \equiv \phi_f$ (water)[b]	ϕ_f (water)[c]
LA27	reverse turn	β-strand$_1$, first and second turns α-helix	−0.03 ± 0.04	0.19 ± 0.01
VA38[d]	α-helix	β-strand$_1$, first and second turns		
IV39	α-helix	β-strand$_1$ and β-strand$_3$	0.52 ± 0.06	0.47 ± 0.03
IV48	β-strand$_2$	α-helix	0.44 ± 0.07	0.29 ± 0.04
IA48	β-strand$_2$	α-helix, β-strand$_3$	0.29 ± 0.03	0.27 ± 0.01
VA66	β-strand$_3$	β-strand$_1$, α-helix	0.28 ± 0.03	0.23 ± 0.01
LA68	β-strand$_3$	α-helix, β-strand$_2$, β-strand$_4$	0.53 ± 0.03	0.55 ± 0.02
VA70	β-strand$_3$	β-strand$_2$, β-strand$_4$, fifth turn	0.44 ± 0.03	0.31 ± 0.02
V76[e]	β-strand$_4$	α-helix, β-strand$_3$		
IA76	β-strand$_4$	α-helix, β-strand$_3$	0.11 ± 0.03	0.10 ± 0.01
IA48/IV76	β-strand$_2$	α-helix, β-strand$_3$	0.34 ± 0.03	0.31 ± 0.01

[a] The interactions listed are not the only ones made by the groups, but they reflect the main interactions (>80%); more specific information is given in the section describing the mutations. [b] ϕ_u is taken from Table I as the values for water. [c] ϕ_f is calculated from $\Delta\Delta G_{\ddagger-U}^{H_2O}$ and $\Delta\Delta G_{U-F}^{[G]_{50\%}}$. [d] ϕ_u and ϕ_f are not calculated for VA38 because $\Delta\Delta G_{\ddagger-F}^{4M} > \Delta\Delta G_{U-F}^{4M}$ and $\Delta\Delta G_{\ddagger-U}^{H_2O} > \Delta\Delta G_{U-F}^{[G]_{50\%}}$ for this mutant. [e] ϕ_u and ϕ_f are not calculated for IV76 because $\Delta\Delta G_{\ddagger-F}$ and $\Delta\Delta G_{U-F}$ are close to 0 for this mutant.

Table IV: Fine-Structure Analysis of Ile → Val and Val and Ile → Ala Mutations

	unfolding				refolding[a]				equilibrium[c]
mutant	m_{k_u} (mol^{-1})	ln $k_u^{H_2O}$	$\Delta\Delta G_{\ddagger-F}^{H_2O}$ (kcal mol^{-1})	$1-\phi_u$ (water)	m_{k_f} (mol^{-1})	ln $k_f^{H_2O}$	$\Delta\Delta G_{\ddagger-U}^{H_2O}$ (kcal mol^{-1})	ϕ_f (water)	$\Delta\Delta G_{U-F}^{[G]_{50\%}}$ (kcal mol^{-1})
Ile → Val48	1.32 ± 0.01	−8.00 ± 0.09	0.61 ± 0.07	0.44 ± 0.07	1.88 ± 0.05	3.49 ± 0.06	0.31 ± 0.04	0.29 ± 0.04	1.09 ± 0.07
Ile → Ala48	1.12 ± 0.03	−4.41 ± 0.12	2.74 ± 0.08	0.29 ± 0.03	2.32 ± 0.10	2.24 ± 0.07	1.05 ± 0.04	0.27 ± 0.01	3.84 ± 0.09
Val → Ala48			2.13 ± 0.11	0.23 ± 0.05			0.74 ± 0.06	0.27 ± 0.02	2.75 ± 0.11
Ile → Val76[b]	1.28 ± 0.03	−9.31 ± 0.15	−0.16 ± 0.10		1.85 ± 0.01	3.75 ± 0.02	0.16 ± 0.02		−0.21 ± 0.10
Ile → Ala76	1.17 ± 0.04	−2.67 ± 0.16	3.77 ± 0.10	0.11 ± 0.03	2.10 ± 0.07	3.30 ± 0.05	0.43 ± 0.03	0.10 ± 0.01	4.25 ± 0.12
Val → Ala76			3.93 ± 0.14	0.12 ± 0.04			0.27 ± 0.04	0.06 ± 0.01	4.46 ± 0.16

[a] Refolding data are combined data from both [GdnHCl]-jump and pH-jump experiments. [b] ϕ_u and ϕ_f are not calculated for IV76 because $\Delta\Delta G_{\ddagger-U}^{H_2O}$ and $\Delta\Delta G_{U-F}^{[G]_{50\%}}$ are close to 0. [c] Data taken from preceding paper (Jackson et al., 1993).

between β-strand$_4$ and β-strand$_3$ and the α-helix. The value ϕ_f for Ile → Ala76 is very close to zero. This shows that the interactions between β-strand$_4$ and β-strand$_3$ and the α-helix are completely unformed in the transition state in this region.

A fine-structure analysis of these two mutations allows one to look at the composite mutation, Val → Ala76 (Table IV). This gives specific information about the interactions that the C$^{\gamma 1}$ and C$^{\gamma 2}$ methyl groups make in the transition state. ϕ_f for Val → Ala76 is again very close to zero. This indicates that all the interactions of Ile76 side chain are not made in the transition state.

(ii) Fractional ϕ_f Values; Interactions Are Partly Formed in the Transition State. (a) Ile → Val39. Ile39 is located in the center of the α-helix. Its side chain is in the center of the hydrophobic core and has many contacts with other side chains in the hydrophobic core. The majority of the interactions that are made by the C$^{\delta 1}$ methyl group are with Trp24 (β-strand$_1$). In addition, this group also interacts with Val66 and Leu68 (β-strand$_3$) and Pro80 (β-strand$_4$). As such, the mutation Ile → Val39 is an excellent probe of the formation of the center of the hydrophobic core and the interaction between the α-helix and the β-sheet (β-strand$_1$ and β-strand$_4$ in particular). ϕ_f is 0.5. This indicates that approximately 50% of the interaction energy made by the C$^{\delta 1}$ methyl group in the native state is maintained in the transition state. This suggests that this group must be in a region which is partially structured in the transition state. There are two possible alternatives: either the C$^{\delta 1}$ methyl group could fully make some interactions but have others completely unformed in the transition state, or all interactions could be somewhat weakened. This suggests that the β-sheet and the center of the α-helix must be formed to some extent in the transition state. It may be that both the β-sheet and the α-helix are formed to a large extent in the transition state but that the two are not packed closely so that interactions between the two are weakened relative to the folded state.

(b) Ile → Val48; Ile → Ala48. Ile48 is the second residue of β-strand$_2$; its side chain is completely buried in the center of the hydrophobic core. The mutation Ile → Val48 removes the C$^{\delta 1}$ methyl group from the side chain and in doing so removes the interactions this group makes with many of the side chains of residues in the α-helix: Lys36, Ile39, and Leu40. Therefore, the mutation Ile → Val48 is an excellent probe for interactions between β-strand$_2$ and the α-helix. ϕ_f is 0.3–0.4. Thus, approximately 60–70% of the interaction energy between β-strand$_2$ and the α-helix has been lost in this region of the core on going from the native state to the transition state in unfolding.

The mutation Ile → Ala48 removes both C$^{\delta 1}$ and C$^{\gamma 1}$ and C$^{\gamma 2}$ methyl groups. This mutation removes further interactions with the α-helix (Lys36 and Ile39), and interactions with β-strand$_3$ (Leu68). In addition, a single interaction is observed with Ala46 (turn$_3$) and Val50 (β-strand$_2$). However, the mutation Ile → Ala48 is still dominated by the loss of interactions between β-strand$_2$ and the α-helix. This is reflected in the value of ϕ_f, which is 0.3.

It is possible to do a fine-structure analysis on this side chain and examine the composite mutation Val → Ala48 (Table IV). This mutation tells us more specifically about the interactions between β-strand$_2$ and the α-helix and β-strand$_3$. In this case, ϕ_f is 0.25. The ϕ_f values for Ile → Val48, Ile → Ala48, and Val → Ala48 are all close, indicating that the interactions made by the different methyl groups of this side chain are all formed to the same degree in the transition state.

(c) Leu → Ala68. Leu68 is the central residue of β-strand$_3$; its side chain is completely buried and in the center of the hydrophobic core of the protein. The C$^{\gamma 1}$, C$^{\delta 1}$, and C$^{\delta 2}$ methyl groups mainly interact with Val32, Ala35, and Ile39 in the α-helix, Ile48 and Val50 in β-strand$_2$, and Ile76 and Pro80 in β-strand$_4$. The side chain of Leu68 makes slightly more interactions within the β-sheet than with the α-helix. There-

Table V: Double-Mutant Cycle Analysis of Ile → Ala48 and Ile → Val76 and the Double-Mutant Ile → Ala48/Ile → Val76 in the Transition State of Folding

	unfolding				refolding[b]	
mutant	$\Delta\Delta G_{\ddagger-F}^{H_2O}$ (kcal mol^{-1})	$\Delta\Delta G_{U-F}^{[G]_{50\%}}$ (kcal mol^{-1})[a]	$\Delta\Delta G_{\ddagger-U}^{H_2O}$ (kcal mol^{-1})	$\Delta\Delta G_{int}$ (kcal mol^{-1})	$\Delta\Delta G_{\ddagger-U}^{H_2O}$ (kcal mol^{-1})	$\Delta\Delta G_{int}$ (kcal mol^{-1})
IA48	2.74 ± 0.08	3.84 ± 0.09	1.10 ± 0.12		1.05 ● 0.04	
IV76	−0.16 ± 0.10	−0.21 ± 0.10	−0.05 ± 0.14		0.16 ± 0.02	
IA48/IV76	2.68 ± 0.09	4.05 ± 0.10	1.37 ± 0.13	0.32 ± 0.23	1.26 ± 0.04	0.05 ± 0.06

[a] Data taken from Jackson et al. (1993). [b] Combined data from [GdnHCl]-jump and pH-jump experiments used in analysis.

fore, Leu → Ala68 is a probe of both β-sheet formation and interaction between β-strand$_3$ and the α-helix. ϕ_f is 0.54, which is the highest value obtained for any of the hydrophobic core mutants. In light of the data obtained for other mutants which suggest that the interaction between the β-sheet and the α-helix is weakened by about 50% in the transition state, this value suggests that the β-sheet formed by β-strands 2, 3, and 4 is present in the transition state in this region of the protein. The φ value therefore reflects the weak interactions between β-strand$_3$ and the α-helix, and not within the β-sheet.

(d) Val → Ala70. Val70 is the last residue in β-strand$_3$. Its side chain interacts with the side chains of residues Val50 (β-strand$_2$), Leu68 (β-strand$_3$), Ile76 (β-strand$_4$), and Val32 (α-helix). However, it also interacts with main-chain atoms of residues Leu51 and Pro52 (β-strand$_2$), Leu68 (β-strand$_3$), Asp74 (turn$_5$), and Asn75, which is the first residue of β-strand$_4$. Mutation of Val → Ala70 gives information on the interaction between β-strand$_3$ and β-strands 2 and 4, and the reverse turn linking β-strand$_3$ and β-strand$_4$. This is a good probe of β-sheet formation in this region of the protein. ϕ_f is 0.3–0.4. Although interactions are very much weakened in the transition state in this region, there is probably still some β-sheet formation in this region of the protein.

(iii) $\phi_f > 1$; the Side Chain Makes More Interactions in the Transition State than in the Native State. Val → Ala38. Val38 is located in the α-helix. It lies right on the edge of the hydrophobic core and has a solvent-accessible surface area of 52 Å^2 (44% of the side-chain surface area). The side chain of Val38 makes only a few contacts in the native state with Trp24 (β-strand$_1$) and Leu27 (reverse turn). This mutant is very unusual in that the difference in the free energy between wild-type and mutant is significantly larger in the transition state than in the native state. The values are 0.73 and 0.86 kcal mol^{-1} for $\Delta\Delta G_{\ddagger-F}^{H_2O}$ and $\Delta\Delta G_{\ddagger-F}^{4M}$, respectively, and 0.46 kcal mol^{-1} for $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$. These results suggest that the side chain of Val38 makes more contacts, some of them nonnative, in the transition state than in the native state. There must, therefore, be structural rearrangements in this region of the protein on going from the transition state to the native state. This result is very unusual and has not been found for any other residue in CI2. It is possible that this is a result of the position of the residue at the very edge of the core and exposed to solvent.

Ile → Ala48/Ile → Val76 (Double Mutant). This mutation is dominated by the contribution from the Ile → Ala48 mutation rather than the Ile → Val76 mutation, that has little effect on the stability either in the native state or in the transition state (Table I). To a first approximation, this mutant should give us similar information, therefore, as the single Ile → Ala48 mutant. Thus, the mutant is a probe of the interaction between β-strand$_2$ and the α-helix and β-strand$_3$. ϕ_f is 0.3 and so is very similar to the values found for Ile → Ala48, indicating that the interaction between β-strand$_2$ and the α-helix and β-strand$_3$ is significantly weakened in the transition state in this region of the protein.

Double-Mutant Cycle Analysis. Double-mutant cycles have been used to measure $\Delta\Delta G_{int}$, the interaction energy between two residues in a protein. The general theory of double-mutant cycles has been discussed extensively elsewhere (Horowitz & Fersht, 1991). $\Delta\Delta G_{int}$ can be used as a probe of structure during protein folding. It is assumed that $\Delta\Delta G_{Uint}$, the interaction energy in the unfolded state, is zero, and this is taken as the reference state. Table V shows the interaction energies between Ile48 and Ile76 measured in the transition state, from both unfolding and refolding data. $\Delta\Delta G_{int}$ measured from refolding data has a significantly lower error than that measured from unfolding data and is close to zero. This indicates that the interaction between Ile48 and Ile76 is not present in the transition state. The interaction must, therefore, form very late on the folding pathway, probably only being realized in the folded state.

DISCUSSION

Validity of Extrapolating Data from Denaturing to Renaturing Conditions. The kinetics of unfolding and refolding of CI2 and its mutants were studied over the complete range of concentrations of GdnHCl. The data fit nicely to a two-state transition in which there is a single rate-determining transition state between folded and unfolded states. The evidence for this is as follows: (i) the simple V-shaped kinetic profiles (Figure 2); (ii) the free energies of unfolding calculated from the ratios of the unfolding and refolding rate constants are in excellent agreement with those measured directly under equilibrium conditions (see Figure 4); (iii) the sum of the kinetic m values, m_{k_u} and m_{k_f}, agrees well with the equilibrium values (see Figure 4) (Jackson et al., 1993).

We were able, therefore, to study directly by kinetics the transition state from both directions of the reaction. Most importantly, we could measure the changes in the energy of the transition state in water in the absence of denaturant from refolding kinetics in water, and then compare these with the values calculated by extrapolating the unfolding kinetics to 0 M denaturant. That is, we can compare φ values measured from folding and unfolding. In almost every case, the values obtained from unfolding experiments are within experimental error of the values measured from refolding experiments. Thus, we have strong evidence that the principle of microscopic reversibility applies to this system and that the unfolding and refolding pathways must be identical. We have shown that the presence of high concentrations of denaturant does not affect the nature of the pathway of folding or the transition state for folding.

φ Value Analysis. The major assumptions of the method are the following: (1) mutation does not affect the pathway of folding; (2) mutation does not significantly change the structure of the folded state; (3) mutation does not perturb the structure of the unfolded state; (4) the target groups do not make new interactions with new partners during the course of the reaction (Fersht et al., 1992). For simple cases such

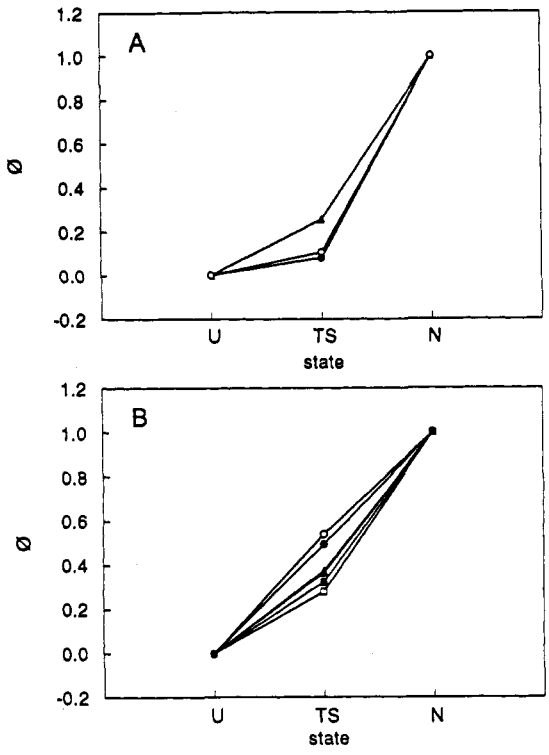

FIGURE 5: ϕ_f plots for the hydrophobic core mutants of CI2. ϕ_f is an average calculated from unfolding data extrapolated to 0 M dentaurant ($1 - \phi_u^{H_2O}$) and refolding data measured directly in water. (A) Residues on the edge of the core: (filled circle) LA27; (open circles) IA76; (filled triangle) VA66. (B) Residues in the center of the core: (filled circle) IV39; (open circle) LA68; (filled triangle) VA70; (open triangle) IV48; (filled squares) IA48IV76; (open square) IA48.

as $\phi_f = 0$ or 1, then assumptions 2 and 3 are not necessarily essential since the effects of the disruption of structure can cancel out (Fersht et al., 1992). All the mutations were designed to minimize the possibility of gross structural rearrangements in the folded protein, and the equilibrium analysis of the mutants in the preceding paper (Jackson et al., 1993) suggests that this is the case. The only exception is the mutation Ile → Val76; structural studies on this protein indicate that there have been small structural rearrangements in the core. However, even in this case the movements of side chains are not large.

We have several probes for each region of the protein, and in some cases for the same position, i.e., Ile → Val and Ile → Ala mutations, which give the same result. Such a consistent pattern of ϕ_f values indicates that the majority of mutations that we have introduced do not produce any major structural change in the folded protein. In addition, this also suggests that the mutant residues do not make different interactions in the transition state from those made in the folded state and that the unfolding/folding pathway has not been significantly changed upon mutation. The single exception is the Val → Ala mutation at position 38. In this case, the residue must (a) make some nonnative interactions in the transition state or (b) change the unfolding/folding pathway. There is no evidence to support the latter; i.e., m_{k_u} and m_{k_f} are very similar to the wild-type values.

Structure of the Transition State. The hydrophobic core mutants studied here yield information about the extent to which the hydrophobic core of CI2 is formed in the transition state for folding. The results are summarized in Table III. The results can be split into two categories: those mutations that result in ϕ_f values close to zero, e.g., LA27, VA66, and IA76; and those mutations that result in fractional ϕ_f values, e.g., IV39, IV48, IA48, LA68, VA70, and IA48/IV76. Plots of the ϕ_f values *versus* the state of the protein for the two groups are shown in Figure 5A,B. Those residues with low ϕ_f values (Figure 5A) are all located on the edge of the hydrophobic core, either in turns or at the beginning or end of a β-strand. Such low ϕ_f values indicate that these residues do not make their interactions in the transition state, and so the edges of the core are not formed in the transition state. In comparison, those residues with ϕ_f values between 0.3 and 0.65 are all located in the center of the core, either in a β-sheet or in a α-helix. Thus, the center of the core is partially formed in the transition state. The interactions are some 50% weaker in the transition state than in the native state, suggesting that the center of the core has not attained the close packing it does in the native state. Interaction energy may be lost in two ways. Nonideal packing would result in a decrease in van der Waals contacts, and may also result in the core becoming partially exposed to solvent, thus reducing the hydrophobic interaction energy.

These results suggest that the breaking of the hydrophobic interactions in the core of the protein is partly rate-determining in unfolding. The last step of refolding, therefore, seems to involve the consolidation of the core of the protein, particularly the edges of the core which appear not to be formed in the transition state.

REFERENCES

Bycroft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488–490.
Chan, H. S., & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6388–6392.
Fresht, A. R., Matouschek, A., & Serrano, L. (1992) *J. Mol. Biol. 224*, 771–782.
Harrison, S. C., & Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4028–4030.
Horovitz, A., Fersht, A. R. (1990) *J. Mol. Biol. 214*, 613–617.
Jackson, S. E., & Fersht, A. R. (1991a) *Biochemistry 30*, 10428–10435.
Jackson, S. E., & Fersht, A. R. (1991b) *Biochemistry 30*, 10428–10435.
Jackson, S. E., Moracci, M., elMasry, N., Johnson, C., & Fersht, A. R. (1993) *Biochemistry* (preceding paper in this issue).
Karplus, M., & Weaver, D. C. (1976) *Nature 260*, 404–4063.
Levinthal, C. (1968) *J. Chem. Phys. 85*, 44–45.
Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1989) *Nature 342*, 122–126.
Matouschek, A., Serrano, L., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 819–835.
Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42048.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 694–699.
Serrano, L., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 805–818.
Staley, J. P., & Kim, P. S. (1990) *Nature 344*, 685–688.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 700–704.