The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-condensation Mechanism for Protein...

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · DECEMBER 1995

Impact Factor: 4.33 · DOI: 10.1006/jmbi.1995.0616 · Source: PubMed

3 AUTHORS:

Laura S Itzhaki
University of Cambridge
77 PUBLICATIONS **3,777** CITATIONS
SEE PROFILE

Daniel Otzen
Aarhus University
237 PUBLICATIONS **8,300** CITATIONS
SEE PROFILE

Alan Fersht
University of Cambridge
629 PUBLICATIONS **52,768** CITATIONS
SEE PROFILE

J. Mol. Biol. (1995) **254**, 260–288

JMB

The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-condensation Mechanism for Protein Folding

Laura S. Itzhaki, Daniel E. Otzen and Alan R. Fersht*

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

The 64-residue protein chymotrypsin inhibitor 2 (CI2) is a single module of structure. It folds and unfolds as a single co-operative unit by simple two-state kinetics *via* a single rate determining transition state. This transition state has been characterized at the level of individual residues by analysis of the rates and equilibria of folding of some 100 mutants strategically distributed at 45 sites throughout the protein. Only one residue, a helical residue (Ala16) buried in the hydrophobic core, has its full native interaction energy in the transition state. The only region of structure which is well developed in the transition state is the α-helix (residues 12 to 24). But, the interactions within it are weakened, especially at the C-terminal region. The rest of the protein has varying degrees of weakly formed structure. Thus, secondary and tertiary interactions appear to form concurrently. These data, reinforced by studies on the structures of peptide fragments, fit a ''nucleation-condensation'' model in which the overall structure condenses around an element of structure, the nucleus, that itself consolidates during the condensation. The high energy transition state is composed of the whole of the molecule making a variety of weak interactions, the nucleus being those residues that make the strongest interactions. The nucleus here is part of the α-helix and some distant residues in the sequence with which it makes contacts. The remainder of the protein has to be sufficiently ordered that it provides the necessary interactions to stabilize the nucleus. The nucleus is only weakly formed in the denatured state but develops in the transition state. The onrush of stability as the nucleus consolidates its local and long range interactions is so rapid that it is not yet fully formed in the transition state. The formation of the nucleus is thus coupled with the condensation. These results are consistent with a recent simulation of the folding of a computer model protein on a lattice which is found to proceed by a nucleation-growth mechanism. We suggest that the mechanism of folding of CI2 may be a common theme in protein folding whereby fundamental folding units of larger proteins, which are modelled by the folding of CI2, form by nucleation-condensation events and coalesce, perhaps in a hierarchical manner.

Keywords: protein folding; transition state; CI2; barnase; protein engineering

**Corresponding author*

Introduction

Protein folding is the process by which a protein progresses from its denatured state to its specific biologically active conformation. It has been proposed that this process has to follow a specific pathway or set of pathways in order to fold in a finite time (Levinthal, 1968), and several different models have been suggested to describe the reaction. According to the hydrophobic collapse model, folding is initiated by burial of hydrophobic

Abbreviations used: CI2, chymotrypsin inhibitor 2; GdmCl, guanidinium chloride; D, denaturant.

0022–2836/95/470260–29 $12.00/0

side-chains (Dill *et al*., 1993). The driving force for folding is visualised as the squeezing out of water from a rapidly formed hydrophobic core within which secondary and tertiary structure is subsequently formed. According to the jigsaw puzzle model (Harrison & Durbin, 1985), however, there is no preferential starting point for folding, and each folding attempt may follow a different path. Recent computer simulations based on lattice models support this idea (Sali *et al*., 1994). Other models envisage pre-formed secondary structural elements diffusing together in collision, and stabilizing each other by local docking rearrangements; the rate-limiting step has been proposed to be variously diffusion (the diffusion-collision-adhesion model; Karplus & Weaver, 1976, 1994) or docking (the framework model; Ptitsyn, 1973, 1991; Kim & Baldwin, 1990) of these secondary elements. These latter models can be distinguished from an early nucleation model (Wetlaufer, 1973), in which the earliest structures are proposed to be formed by a slow random search followed by rapid growth and coalescence into the native folded structure.

Ultimately, the folding pathway of a particular protein is defined when all conformational states of the pathway, including the denatured state, possible intermediates, the major transition state and the native state, as well as their rates of interconversion, are known in detail at the level of individual residues (Fersht, 1993). Owing to the small difference in stability between the folded and unfolded form of a protein and the crudity of currently available potential energy functions, it is difficult to use computer calculations to predict the stability of a specific protein conformation, let alone the feasibility of different pathways, with confidence. It is necessary, therefore, to gather a large experimental database on the stabilizing features of protein structure, and to monitor these interactions, in structural and energetic terms, during folding. This laboratory employs a procedure (''the protein engineering method'') for studying the structure of transition state and unstable intermediates at the level of individual residues by making kinetic measurements on the folding and unfolding of mutant proteins, and relating the changes in rate constants to the changes in equilibrium constants on mutation (Matouschek *et al*., 1989, 1992; Fersht *et al*., 1992; Serrano *et al*., 1992a; Fersht, 1993, 1995a). The principle of the method is that the change in stability of the protein on mutation ($\Delta\Delta G_{F\text{-}U}$) is measured, as is also the change in stability of the transition state for folding ($\Delta\Delta G_{‡\text{-}U}$), and the two values are compared. If the region of the protein at the site of mutation is as folded in the transition state as in the final folded structure, then $\Delta\Delta G_{‡\text{-}U} = \Delta\Delta G_{F\text{-}U}$. If that region is as unfolded as in the denatured protein, then $\Delta\Delta G_{‡\text{-}U} = 0$. The ratio of the two energies, $\Delta\Delta G_{‡\text{-}U} / \Delta\Delta G_{F\text{-}U}$, is defined as Φ_F, which varies from 0 for the example of being completely denatured in the transition state to 1.0 for being completely folded. The large scale application of this procedure to a protein is a major undertaking, which can be simplified by working with small monomeric proteins, especially those that have minimal residual structure in the unfolded state. The method has been employed in depth to the analysis of barnase, a small ribonuclease, which folds *via* a metastable folding intermediate (reviewed by Fersht, 1993).

Here, we present a detailed description of the transition state for folding of the 64-residue monomeric protein chymotrypsin inhibitor 2 (CI2). The folding of CI2 is simplified by the lack of disulphide bridges and *cis*-peptidyl-prolyl bonds in the native state. Its native structure is known in the crystal state (McPhalen & James, 1987; Harpaz *et al*., 1994) and in solution (Ludvigsen *et al.*, 1991). The protein consists of a single domain (or module) that does not have readily discernible substructures that make interactions primarily within themselves. Previous studies have established that the folding and unfolding of wild-type CI2 and a range of mutants conform to a single two-state model under both equilibrium and non-equilibrium conditions (Jackson & Fersht, 1991a,b; Jackson *et al*., 1993a,b). In particular, the ratio of rate constants for unfolding and refolding give the correct equilibrium constant for unfolding (Jackson & Fersht, 1991a; Jackson *et al*., 1993b). This does not mean that there are no folding intermediates in the pathway but simply that any such intermediates are not significantly populated at equilibrium or during the approach to equilibrium. There is thus a single rate-determining transition state for both unfolding and refolding (for a discussion of transition states in protein folding see Fersht, 1995a). A thermodynamic characterization of the transition state for folding/unfolding of CI2 indicates that 75% of the total decrease in heat capacity, ΔC_P, between denatured and native states occurs on going from the denatured to the transition state, and this indicates a significant burial of hydrophobic side-chains in the transition state (Jackson & Fersht, 1991b). This is consistent with the large positive enthalpy of activation of folding at lower temperatures. The occurrence of simple two-state kinetics has the important consequence that the single rate-determining transition state can be studied in the direction of both refolding and unfolding.

A preliminary description of this transition state, derived from the protein engineering method, showed a structure that is like an expanded form of the folded structure in which most interactions in the protein have been greatly weakened (Otzen *et al*., 1994). The major element of structure that is best, but not completely, formed energetically is the single α-helix. This contrasts with the structure for the major transition state for the folding of barnase, which has many elements of structure either fully formed or fully denatured (Fersht, 1993). It was proposed that the folding of CI2 is representative of the folding of a single module protein and is, perhaps, a model for the folding of a domain

(module) in a larger protein. The small size of CI2 makes it an attractive target for computer simulations, and the first such work on the transition state for folding/unfolding gives results in agreement with experiment (Li & Daggett, 1994, and unpublished data). Since it is likely that the folding of CI2 will be used to benchmark and refine continuing theoretical studies on protein folding, it is necessary to define the structure of the transition state of CI2 with precision and present it in detail. It is important to obtain as much data as possible for this process, especially as any one single mutation does have the possibility of intrusion from artifacts. We now report an extensive set of experimental data, including double mutant cycles and the recently introduced Ala → Gly scanning procedure (Matthews & Fersht, 1995), to delineate in detail the transition state for the folding of the barley chymotrypsin inhibitor, analysed from both folding and unfolding kinetics. The structure of the transition state points towards a mechanism for the folding, namely nucleation-condensation.

Results

CI2 has been shown to follow the two-state model of protein folding, under both equilibrium and non-equilibrium conditions (Jackson & Fersht, 1991a; Jackson *et al*., 1993b), i.e. it has only one kinetically significant transition state. Data analysis has been described in detail previously (Jackson *et al*., 1993a,b), and is briefly summarised here.

GdmCl equilibrium denaturation

It is usually assumed that there is a linear relationship between the free energy of unfolding in the presence of GdmCl (abbreviated here to D) and the concentration of denaturant (Tanford, 1968; Pace, 1986):

$$\Delta G^{D}_{U\text{-}F} = \Delta G^{H_2O}_{U\text{-}F} - m_{U\text{-}F}[\text{denaturant}] \qquad (1)$$

where $\Delta G^{D}_{U\text{-}F}$ is the free energy of unfolding at a particular denaturant concentration, D, $\Delta G^{H_2O}_{U\text{-}F}$ is the free energy of unfolding in water, and $m_{U\text{-}F}$ is a constant that is proportional to the increase in the degree of exposure of the protein on denaturation. From equation (1), it is apparent that at $[D]_{50\%}$, the concentration of GdmCl at which 50% of the protein is denatured, $\Delta G^{H_2O}_{U\text{-}F} = m_{U\text{-}F}[D]_{50\%}$; thus:

$$\Delta G^{D}_{U\text{-}F} = m_{U\text{-}F}([D]_{50\%} - [D]) \qquad (2)$$

The denaturation curves were fitted to an equation, derived from (2) above (Clarke & Fersht, 1993), which yields the values for $[D]_{50\%}$ and $m_{U\text{-}F}$ and their standard errors (Table 1). All errors are calculated from the best fit of the data and are not standard errors from repetitive runs, unless otherwise stated. Values of $\Delta G^{H_2O}_{U\text{-}F}$ are also given in Table 1.

Repetitive measurements of $m_{U\text{-}F}$ for an individual mutant have a variability of ± 5 to 10%, whereas $[D]_{50\%}$ is very reproducible at ± 0.02 M because $[D]_{50\%}$ is insensitive to small errors in baselines. It is clear that the $m_{U\text{-}F}$-values for wild-type and mutants are the same within experimental error, apart from a few outliers, which are statistically expected in a large data base. Therefore, the value of $\Delta\Delta G^{[D]_{50\%}}_{U\text{-}F}$, the difference in the free energy of unfolding between wild-type and mutant proteins at a mean value of the $[D]_{50\%}$ for the two proteins, can be calculated from the equation:

$$\Delta\Delta G^{[D]_{50\%}}_{U\text{-}F} = \langle m_{U\text{-}F} \rangle \Delta[D]_{50\%} \qquad (3)$$

where $\langle m_{U\text{-}F} \rangle$ is the average value of m, obtained from measurements on all the mutant proteins of $1.90(\pm 0.03)$ kcal/mol^{-1} M^{-1} ($n = 124$, including unpublished mutants). The use of a mean value of $m_{U\text{-}F}$ allows calculation of the change in the free energy of unfolding on mutation with a low standard error. Equation (3) is very robust since it gives accurate values of $\Delta\Delta G^{[D]_{50\%}}_{U\text{-}F}$ even if the linearity assumption of equation (1) breaks down but there is a small curvature in the plots that is the same for each mutant (Matouschek *et al*., 1994).

A value of $\Delta\Delta G_{U\text{-}F}$ can also be calculated using the individual m values:

$$\Delta\Delta G^{H_2O}_{U\text{-}F} = \Delta G^{H_2O}_{U\text{-}F} - \Delta G'^{H_2O}_{U\text{-}F} \qquad (4)$$

where $\Delta G^{H_2O}_{U\text{-}F}$ and $\Delta G'^{H_2O}_{U\text{-}F}$ are the free energies of unfolding in water for wild-type and mutant, respectively, or at any denaturant concentration using the more general equation:

$$\Delta\Delta G^{[D]}_{U\text{-}F} = m_{U\text{-}F}([D]_{50\%} - [D]) - m'_{U\text{-}F}([D]'_{50\%} - [D]) \qquad (5)$$

where $m'_{U\text{-}F}$ and $[D]'_{50\%}$ are the m-value and midpoint of denaturation, respectively, for the mutant protein, and $m_{U\text{-}F}$ and $[D]_{50\%}$ are the values for wild-type (Clarke & Fersht, 1993). The relative merits of these three methods for calculating $\Delta\Delta G_{U\text{-}F}$ and the errors of the resulting values have been discussed in detail previously (Jackson *et al*., 1993a). $\Delta\Delta G^{[D]_{50\%}}_{U\text{-}F}$, however, has the smallest errors and, for mutations in the hydrophobic core of CI2, shows the best agreement with $\Delta\Delta G_{U\text{-}F}$ measured directly in water by calorimetry (Jackson *et al*., 1993a). Values of $\Delta\Delta G^{[D]_{50\%}}_{U\text{-}F}$ are given in Table 1. Values of $\Delta G^{H_2O}_{U\text{-}F}$ have not been given because the long extrapolation results in very large errors associated with this parameter.

Unfolding kinetics

Plots of the natural logarithm of the rate constants of unfolding against the final GdmCl concentration are linear, within experimental error over the experimentally accessible range of concentrations, conforming to the equation:

$$\ln k_u = \ln k_u^{H_2O} + m_{ku}[D] \qquad (6)$$

where k_u is the rate constant of unfolding at a given GdmCl concentration, $k_u^{H_2O}$ is the rate constant of unfolding in water, m_{ku} is the slope, and [D] is the

Table 1. changes in the free energies of unfolding upon mutation determined by reversible guanidinium chloride denaturation experiments

Mutant	$m_{U\text{-}F}$ (kcal mol^{-1} M^{-1})	$[D]_{50\%}$ (M)	$\Delta G^{H_2O}_{U\text{-}F}$[a] (kcal mol^{-1})	$\Delta\Delta G^{[D]_{50\%}}_{U\text{-}F}$[b] (kcal mol^{-1})
Wild-type	1.90 ± 0.03	4.00 ± 0.01	7.60 ± 0.12	0
KA2	1.77 ± 0.05	3.72 ± 0.01	6.56 ± 0.18	0.55 ± 0.04
KA2/EA7	1.87 ± 0.07	3.43 ± 0.02	6.43 ± 0.24	1.10 ± 0.04
KM2	1.88 ± 0.05	3.66 ± 0.01	6.87 ± 0.19	0.67 ± 0.03
TA3[e]	1.83 ± 0.09	3.56 ± 0.02	6.51 ± 0.31	0.85 ± 0.05
TV3[e]	1.67 ± 0.08	3.83 ± 0.03	6.39 ± 0.32	0.32 ± 0.07
TG3[e]	1.95 ± 0.12	3.40 ± 0.03	6.65 ± 0.41	1.16 ± 0.06
PA6	3.46 ± 0.75	3.19 ± 0.05	11.0 ± 2.40	1.57 ± 0.10
PA6/AG16	2.39 ± 0.14	2.64 ± 0.02	6.30 ± 0.37	2.65 ± 0.06
EA7	1.75 ± 0.06	3.76 ± 0.02	6.57 ± 0.24	0.47 ± 0.05
EQ7	1.91 ± 0.12	3.68 ± 0.03	7.03 ± 0.44	0.62 ± 0.07
LA8[c]	2.15 ± 0.13	2.62 ± 0.02	5.63 ± 0.34	2.68 ± 0.06
KA11	1.52 ± 0.08	4.22 ± 0.02	6.40 ± 0.34	–0.42 ± 0.05
SG12[d]	2.10 ± 0.07	3.59 ± 0.02	7.54 ± 0.25	0.80 ± 0.05
SA12[d]	1.86 ± 0.08	3.54 ± 0.02	6.58 ± 0.29	0.89 ± 0.05
EQ14[d]	1.87 ± 0.10	3.85 ± 0.03	7.20 ± 0.39	0.29 ± 0.06
ED14[d]	1.83 ± 0.11	3.73 ± 0.04	6.83 ± 0.42	0.52 ± 0.08
EN14[d]	2.01 ± 0.10	3.64 ± 0.02	7.32 ± 0.37	0.70 ± 0.05
EQ15[d]	1.84 ± 0.09	3.76 ± 0.03	6.92 ± 0.34	0.47 ± 0.06
ED15[d]	2.20 ± 0.16	3.62 ± 0.03	7.96 ± 0.58	0.74 ± 0.06
EN15[d]	2.18 ± 0.11	3.45 ± 0.02	7.52 ± 0.38	1.07 ± 0.05
EA14/EA15[d]	2.11 ± 0.18	3.61 ± 0.03	7.62 ± 0.65	0.76 ± 0.06
SG12/EA14/EA15[d]	2.15 ± 0.16	3.16 ± 0.03	6.79 ± 0.51	1.63 ± 0.07
SA12/EA14/EA15[d]	2.18 ± 0.14	3.14 ± 0.02	6.85 ± 0.44	1.67 ± 0.05
AG16	1.80 ± 0.07	3.44 ± 0.02	6.19 ± 0.25	1.09 ± 0.05
KA17	1.73 ± 0.05	3.75 ± 0.01	6.47 ± 0.17	0.49 ± 0.03
KG17	1.99 ± 0.10	2.80 ± 0.02	5.59 ± 0.27	2.32 ± 0.06
KA18	1.61 ± 0.13	4.11 ± 0.07	6.61 ± 0.53	–0.21 ± 0.13
KG18	1.73 ± 0.13	3.49 ± 0.04	6.02 ± 0.48	0.99 ± 0.08
VA19[d]	2.01 ± 0.15	3.75 ± 0.03	7.54 ± 0.57	0.49 ± 0.06
IV20[d]	1.99 ± 0.12	3.33 ± 0.02	6.63 ± 0.40	1.30 ± 0.05
LA21	1.93 ± 0.15	3.32 ± 0.03	6.39 ± 0.49	1.33 ± 0.07
LG21	1.68 ± 0.07	3.29 ± 0.02	5.51 ± 0.23	1.38 ± 0.05
QA22	1.77 ± 0.15	3.99 ± 0.05	7.06 ± 0.60	0.02 ± 0.10
QG22	1.81 ± 0.12	3.69 ± 0.03	6.68 ± 0.43	0.60 ± 0.06
DA23	1.87 ± 0.05	3.51 ± 0.01	6.55 ± 0.16	0.96 ± 0.03
KA24	1.57 ± 0.08	3.67 ± 0.03	5.75 ± 0.28	0.65 ± 0.06
KG24	1.75 ± 0.17	2.36 ± 0.05	4.11 ± 0.40	3.19 ± 0.11
PA25	2.07 ± 0.10	3.09 ± 0.02	6.38 ± 0.32	1.76 ± 0.05
EA26	1.78 ± 0.06	3.83 ± 0.01	6.82 ± 0.24	0.32 ± 0.03
EM28/ML40	1.78 ± 0.05	4.16 ± 0.02	7.42 ± 0.21	–0.32 ± 0.05
IV29[c]	1.99 ± 0.13	3.43 ± 0.02	6.83 ± 0.45	1.11 ± 0.05
IA29[c]	2.07 ± 0.19	1.99 ± 0.03	4.12 ± 0.38	3.90 ± 0.09
IA29/IV57[c]	1.83 ± 0.10	1.90 ± 0.02	3.48 ± 0.19	4.08 ± 0.08
IV30[e]	1.77 ± 0.05	4.04 ± 0.04	7.17 ± 0.19	–0.08 ± 0.08
IA30[e]	2.22 ± 0.09	2.91 ± 0.02	6.45 ± 0.27	2.12 ± 0.06
IG30[e]	2.13 ± 0.08	2.19 ± 0.01	4.65 ± 0.17	3.52 ± 0.07
IT30[e]	1.97 ± 0.07	3.31 ± 0.02	6.53 ± 0.24	1.34 ± 0.04
LA32	2.11 ± 0.09	2.78 ± 0.02	5.86 ± 0.27	2.37 ± 0.05
LI32	1.70 ± 0.08	3.87 ± 0.03	6.59 ± 0.33	0.26 ± 0.07
LV32	1.76 ± 0.10	3.74 ± 0.03	6.58 ± 0.36	0.50 ± 0.06
LV32/FL50	2.37 ± 0.13	2.75 ± 0.02	6.52 ± 0.37	2.42 ± 0.06
LV32/FA50	1.94 ± 0.10	2.24 ± 0.02	4.34 ± 0.22	3.41 ± 0.07
LA32/FL50	1.94 ± 0.09	2.24 ± 0.02	4.35 ± 0.19	3.42 ± 0.07
LA32/FA50	1.59 ± 0.14	1.53 ± 0.08	2.44 ± 0.25	4.79 ± 0.19
LA32/VA38	1.96 ± 0.06	2.37 ± 0.01	4.65 ± 0.14	3.16 ± 0.06
LV32/VA38	1.86 ± 0.06	3.05 ± 0.01	5.67 ± 0.20	1.85 ± 0.05
LA32/VA38/FL50	1.93 ± 0.06	2.21 ± 0.01	4.26 ± 0.13	3.48 ± 0.07
LV32/VA38/FL50	1.83 ± 0.05	2.60 ± 0.01	4.76 ± 0.13	2.72 ± 0.06
PA33	1.79 ± 0.07	3.91 ± 0.02	7.02 ± 0.27	0.17 ± 0.05
VT34[e]	1.65 ± 0.08	3.47 ± 0.02	5.74 ± 0.26	1.03 ± 0.05
VA34[e]	1.70 ± 0.15	3.67 ± 0.05	6.24 ± 0.57	0.64 ± 0.11
VG34[e]	2.00 ± 0.08	2.75 ± 0.01	5.50 ± 0.22	2.43 ± 0.05
TS36	1.66 ± 0.03	3.99 ± 0.01	6.62 ± 0.13	0.02 ± 0.02
TV36	1.67 ± 0.12	3.61 ± 0.04	6.03 ± 0.43	0.76 ± 0.08
TA36	1.50 ± 0.05	4.12 ± 0.01	6.18 ± 0.21	–0.23 ± 0.03
IA37	1.59 ± 0.06	3.98 ± 0.03	6.32 ± 0.25	0.03 ± 0.06
IA37Δ38	1.90 ± 0.05	2.52 ± 0.01	4.78 ± 0.13	2.88 ± 0.06
VA38	2.14 ± 0.12	3.24 ± 0.02	6.93 ± 0.38	1.47 ± 0.05

continued

Table 1. *continued*

Mutant	$m_{U\text{-}F}$ (kcal mol^{-1} M^{-1})	$[D]_{50\%}$ (M)	$\Delta G_{U\text{-}F}^{H_2O}$[a] (kcal mol^{-1})	$\Delta\Delta G_{U\text{-}F}^{[D]_{50\%}}$[b] (kcal mol^{-1})
VA38/FL50	2.02 ± 0.08	2.67 ± 0.01	5.40 ± 0.20	2.58 ± 0.06
VA38/FA60	2.19 ± 0.12	1.75 ± 0.02	3.82 ± 0.21	4.37 ± 0.09
TA39[f]	1.89 ± 0.14	3.63 ± 0.04	6.86 ± 0.51	0.72 ± 0.08
TD39[f]	1.87 ± 0.08	4.01 ± 0.03	7.50 ± 0.33	−0.02 ± 0.06
TA39/EA41[f]	1.94 ± 0.10	3.54 ± 0.02	6.87 ± 0.36	0.89 ± 0.05
TD39/EA41[f]	1.89 ± 0.09	3.86 ± 0.02	7.30 ± 0.35	0.27 ± 0.04
EA41	1.78 ± 0.16	3.64 ± 0.05	6.48 ± 0.59	0.70 ± 0.10
RA43	1.78 ± 0.10	3.70 ± 0.03	6.58 ± 0.38	0.58 ± 0.07
RA43/DA45	1.53 ± 0.10	3.37 ± 0.04	5.14 ± 0.36	1.22 ± 0.08
DA45	1.83 ± 0.08	3.59 ± 0.02	6.58 ± 0.28	0.80 ± 0.05
VA47[c]	1.76 ± 0.22	1.46 ± 0.10	2.57 ± 0.37	4.93 ± 0.21
LA49[c]	1.98 ± 0.12	2.02 ± 0.03	4.00 ± 0.25	3.84 ± 0.09
FL50	1.95 ± 0.11	2.91 ± 0.02	5.68 ± 0.33	2.11 ± 0.06
FV50	2.24 ± 0.18	2.77 ± 0.03	6.19 ± 0.51	2.39 ± 0.07
FA50	2.17 ± 0.13	2.02 ± 0.03	4.39 ± 0.26	3.84 ± 0.08
VA51[c]	2.15 ± 0.19	2.98 ± 0.03	6.41 ± 0.57	1.98 ± 0.07
DA52	1.99 ± 0.21	2.24 ± 0.04	4.46 ± 0.47	3.41 ± 0.10
DN53	1.60 ± 0.03	4.00 ± 0.01	6.39 ± 0.12	−0.004 ± 0.03
ND56	1.80 ± 0.07	3.38 ± 0.02	6.09 ± 0.24	1.21 ± 0.05
NA56	1.79 ± 0.04	3.57 ± 0.01	6.38 ± 0.14	0.83 ± 0.03
IV57[c]	1.82 ± 0.12	4.10 ± 0.05	7.46 ± 0.50	−0.19 ± 0.10
IA57[c]	1.93 ± 0.21	1.79 ± 0.05	3.45 ± 0.39	4.29 ± 0.12
AG58[e]	2.19 ± 0.22	3.03 ± 0.04	6.65 ± 0.67	1.88 ± 0.08
VT60[e]	1.63 ± 0.14	3.80 ± 0.06	6.20 ± 0.53	0.38 ± 0.11
VA60[e]	1.92 ± 0.11	3.22 ± 0.03	6.17 ± 0.37	1.51 ± 0.06
VG60[e]	2.61 ± 0.11	2.33 ± 0.01	6.08 ± 0.27	3.24 ± 0.06
PA61	1.80 ± 0.14	2.28 ± 0.04	4.10 ± 0.32	3.34 ± 0.09
VT63[e]	1.79 ± 0.13	3.41 ± 0.03	6.08 ± 0.45	1.15 ± 0.07
VA63[e]	1.95 ± 0.14	3.25 ± 0.03	6.33 ± 0.45	1.45 ± 0.07
VG63[e]	2.10 ± 0.06	2.20 ± 0.01	4.61 ± 0.14	3.50 ± 0.07

Measured at 25°C, 50 mM Mes (pH 6.25).

[a] $\Delta G_{U\text{-}F}^{H_2O} = m_{U\text{-}F}[D]_{50\%}$.

[b] $\Delta\Delta G_{U\text{-}F}^{[D]_{50\%}} = \langle m_{U\text{-}F}\rangle([D]_{50\%} - [D]'_{50\%})$, where $\langle m_{U\text{-}F}\rangle$ is the mean value of $m_{U\text{-}F}$, from measurements on all the mutant proteins and repetitive runs on wild-type, of 1.94(±0.033) kcal mol^{-2}, and $[D]'_{50\%}$ and $[D]_{50\%}$ are the concentrations of GdmCl at which 50% of mutant protein and wild-type protein, respectively, are denatured. Note that $\Delta\Delta G_{U\text{-}F}^{[D]_{50\%}}$ is defined as $\Delta G_{U\text{-}F}$ (wild type) − $\Delta G_{U\text{-}F}$ (mutant).

[c] From Jackson *et al*. (1993a).

[d] From elMasry & Fersht (1994).

[e] From Otzen & Fersht (1995).

[f] From Jackson & Fersht (1994).

GdmCl concentration. If ln k_u is plotted against [D], then the plots can be extrapolated to obtain ln k_u^{4M}. ln $k_u^{H_2O}$, ln k_u^{4M} and m_{ku} for all mutants are given in Table 2. The standard errors are significantly lower for ln k_u^{4M} compared with ln $k_u^{H_2O}$ because of the shorter extrapolation.

The stability of the transition state of the mutant protein relative to that of the wild-type protein is calculated from the unfolding kinetics from the equation:

$$\Delta\Delta G_{\ddagger\text{-}F} = -RT \ln(k_u / k'_u) \quad (7)$$

where $\Delta\Delta G_{\ddagger\text{-}F}$ is the difference in energy of the transition state of unfolding relative to the folded state between wild-type and mutant, and k_u and k'_u are the rate constants of unfolding for the wild-type and mutant, respectively. The values of $\Delta\Delta G_{\ddagger\text{-}F}$ are given in Table 2.

Refolding kinetics

Jackson & Fersht (1991b) have shown that there is a series of slow refolding phases that arise from *cis-trans* isomerisation of peptidyl-prolyl bonds. The analysis of energetics and pathway in this study concerns only the major fast refolding phase of the all-*trans* peptidyl-proline from of the unfolded state to the all-*trans* folded state that accounts for approximately 70% of the total amplitude.

Plots of the natural logarithm of the rate constants of folding against the final GdmCl concentration are linear, conforming to the equation:

$$\ln k_f = \ln k_f^{H_2O} + m_{kf}[D] \quad (8)$$

where k_f is the rate constant of folding at a given GdmCl concentration, $k_f^{H_2O}$ is the rate constant of folding in water, m_{kf} is the slope, and [D] is the final GdmCl concentration. ln $k_f^{H_2O}$ and m_{kf} for all mutants are given in Table 3.

The stability of the transition state of the mutant protein relative to that of the wild-type protein is calculated from the folding kinetics, in a similar manner to the analysis of the unfolding data, using the equation:

$$\Delta\Delta G_{\ddagger\text{-}U} = -RT \ln(k_f / k'_f) \quad (9)$$

where $\Delta\Delta G_{\ddagger\text{-}U}$ is the difference in energy of the transition state of unfolding relative to the folded

Table 2. Rates and thermodynamic data from unfolding kinetics of CI2 mutants

Mutant	$\ln k_u^{H_2O}$	$\ln k_u^{4M}$	m_{ku} (M^{-1})	$\Delta\Delta G_{‡\text{-}F}^{H_2O}$[a] (kcal mol^{-1})	$\Delta\Delta G_{‡\text{-}F}^{4M}$[a] (kcal mol^{-1})
Wild type	–9.04 ± 0.07	–3.90 ± 0.02	1.31 ± 0.01	0	0
KA2	–7.83 ± 0.08	–2.72 ± 0.03	1.28 ± 0.01	–0.72 ± 0.06	–0.67 ± 0.02
KA2/EA7	–7.36 ± 0.22	–2.36 ± 0.07	1.25 ± 0.04	–1.00 ± 0.14	–0.91 ± 0.04
KA2/DA23	–6.74 ± 0.13	–1.57 ± 0.04	1.29 ± 0.02	–1.36 ± 0.09	–1.38 ± 0.02
KM2	–8.25 ± 0.13	–2.93 ± 0.04	1.33 ± 0.02	–0.47 ± 0.09	–0.57 ± 0.03
TA3	–7.98 ± 0.10	–2.73 ± 0.03	1.31 ± 0.02	–0.63 ± 0.07	–0.69 ± 0.02
TV3	–8.58 ± 0.06	–3.41 ± 0.02	1.29 ± 0.01	–0.27 ± 0.05	–0.29 ± 0.01
TG3	–7.68 ± 0.16	–2.51 ± 0.05	1.29 ± 0.03	–0.81 ± 0.10	–0.83 ± 0.03
PA6	–7.62 ± 0.49	–1.24 ± 0.11	1.60 ± 0.10	–0.84 ± 0.29	–1.58 ± 0.07
PA6/AG16	–6.65 ± 0.06	–1.28 ± 0.02	1.34 ± 0.01	–1.42 ± 0.05	–1.55 ± 0.01
EA7	–8.59 ± 0.13	–3.39 ± 0.05	1.30 ± 0.02	–0.27 ± 0.09	–0.30 ± 0.03
LA8[b]	–4.42 ± 0.14	–0.12 ± 0.03	1.05 ± 0.03	–2.73 ± 0.09	–2.24 ± 0.02
KA11	–9.39 ± 0.05	–4.32 ± 0.02	1.27 ± 0.01	0.21 ± 0.05	0.24 ± 0.01
SG12	–8.42 ± 0.22	–3.24 ± 0.08	1.29 ± 0.04	–0.36 ± 0.14	–0.39 ± 0.05
SA12	–8.53 ± 0.22	–3.37 ± 0.08	1.29 ± 0.04	–0.30 ± 0.14	–0.32 ± 0.05
EQ14	–8.97 ± 0.12	–3.70 ± 0.04	1.32 ± 0.02	–0.04 ± 0.08	–0.12 ± 0.03
ED14	–8.76 ± 0.29	–3.43 ± 0.09	1.33 ± 0.05	–0.16 ± 0.18	–0.28 ± 0.05
EN14	–9.01 ± 0.13	–3.51 ± 0.05	1.38 ± 0.02	–0.02 ± 0.09	–0.23 ± 0.03
EQ15	–8.71 ± 0.16	–3.53 ± 0.05	1.30 ± 0.03	–0.20 ± 0.10	–0.22 ± 0.03
ED15	–8.28 ± 0.08	–2.82 ± 0.03	1.37 ± 0.01	–0.45 ± 0.06	–0.64 ± 0.02
EN15	–7.95 ± 0.10	–2.89 ± 0.06	1.26 ± 0.02	–0.65 ± 0.07	–0.60 ± 0.04
EA14/EA15	–8.17 ± 0.23	–2.95 ± 0.08	1.31 ± 0.04	–0.52 ± 0.14	–0.56 ± 0.05
SG12/EA14/EA15	–6.66 ± 0.28	–1.95 ± 0.09	1.18 ± 0.05	–1.41 ± 0.17	–1.16 ± 0.05
SA12/EA14/EA15	–7.03 ± 0.10	–2.18 ± 0.03	1.21 ± 0.02	–1.19 ± 0.07	–1.02 ± 0.02
AG16	–9.47 ± 0.14	–4.24 ± 0.05	1.31 ± 0.02	0.26 ± 0.09	0.20 ± 0.03
KA17	–7.83 ± 0.12	–2.87 ± 0.04	1.24 ± 0.02	–0.72 ± 0.08	–0.61 ± 0.03
KG17	–6.13 ± 0.11	–1.63 ± 0.03	1.13 ± 0.02	–1.72 ± 0.08	–1.35 ± 0.02
KA18	–8.89 ± 0.13	–3.64 ± 0.05	1.31 ± 0.02	–0.09 ± 0.09	–0.15 ± 0.03
KG18	–8.35 ± 0.11	–3.13 ± 0.04	1.30 ± 0.02	–0.41 ± 0.08	–0.46 ± 0.03
VA19[b]	–7.80 ± 0.13	–2.45 ± 0.05	1.35 ± 0.02	–0.73 ± 0.09	–0.86 ± 0.03
IV20[b]	–8.02 ± 0.11	–2.94 ± 0.03	1.27 ± 0.02	–0.60 ± 0.08	–0.57 ± 0.02
LA21	–7.44 ± 0.09	–2.77 ± 0.02	1.17 ± 0.01	–0.95 ± 0.07	–0.67 ± 0.01
LG21	–7.19 ± 0.04	–2.61 ± 0.01	1.14 ± 0.01	–1.10 ± 0.05	–0.76 ± 0.01
QA22	–8.98 ± 0.06	–3.72 ± 0.02	1.32 ± 0.01	–0.03 ± 0.06	–0.11 ± 0.02
QG22	–8.27 ± 0.09	–3.21 ± 0.03	1.27 ± 0.01	–0.46 ± 0.07	–0.41 ± 0.02
DA23	–7.87 ± 0.21	–2.36 ± 0.06	1.38 ± 0.04	–0.69 ± 0.13	–0.91 ± 0.04
KA24	–7.90 ± 0.06	–2.72 ± 0.02	1.30 ± 0.01	–0.68 ± 0.06	–0.70 ± 0.02
KG24	–4.48 ± 0.10	1.00 ± 0.01	1.37 ± 0.03	–2.70 ± 0.07	–2.90 ± 0.01
PA25	–6.43 ± 0.19	–1.78 ± 0.06	1.16 ± 0.03	–1.55 ± 0.12	–1.25 ± 0.04
EA26	–8.61 ± 0.09	–3.49 ± 0.03	1.28 ± 0.01	–0.25 ± 0.07	–0.25 ± 0.02
IV29[b]	–8.00 ± 0.09	–2.73 ± 0.03	1.32 ± 0.01	–0.62 ± 0.07	–0.69 ± 0.02
IA29[b]	–4.30 ± 0.18	0.10 ± 0.03	1.10 ± 0.04	–2.80 ± 0.11	–2.37 ± 0.02
IA29/IV57[b]	–4.33 ± 0.15	0.04 ± 0.04	1.09 ± 0.03	–2.79 ± 0.10	–2.33 ± 0.03
IV30	–8.99 ± 0.13	–4.13 ± 0.05	1.22 ± 0.02	–0.03 ± 0.09	0.13 ± 0.03
IA30	–7.50 ± 0.08	–2.52 ± 0.03	1.25 ± 0.01	–0.91 ± 0.06	–0.82 ± 0.02
IG30	–5.29 ± 0.13	–0.19 ± 0.04	1.28 ± 0.03	–2.22 ± 0.09	–2.20 ± 0.02
IT30	–8.19 ± 0.09	–3.25 ± 0.03	1.24 ± 0.02	–0.50 ± 0.07	–0.39 ± 0.02
LA32	–6.34 ± 0.23	–1.34 ± 0.06	1.25 ± 0.04	–1.60 ± 0.14	–1.51 ± 0.04
LI32	–8.54 ± 0.03	–3.32 ± 0.01	1.31 ± 0.01	–0.29 ± 0.05	–0.35 ± 0.01
LV32	–8.22 ± 0.04	–2.96 ± 0.01	1.31 ± 0.01	–0.48 ± 0.05	–0.56 ± 0.01
LV32/FL50	–6.17 ± 0.04	–1.19 ± 0.01	1.25 ± 0.01	–1.70 ± 0.05	–1.61 ± 0.01
LV32/FA50	–5.08 ± 0.11	–0.52 ± 0.03	1.14 ± 0.02	–2.34 ± 0.08	–2.00 ± 0.02
LA32/FL50	–5.07 ± 0.07	–0.56 ± 0.02	1.13 ± 0.01	–2.35 ± 0.06	–1.98 ± 0.01
LA32/FA50	–3.65 ± 0.10	0.55 ± 0.03	1.05 ± 0.02	–3.19 ± 0.07	–2.64 ± 0.02
LA32/VA38	–5.34 ± 0.10	–0.52 ± 0.02	1.20 ± 0.02	–2.19 ± 0.07	–2.00 ± 0.02
LV32/VA38	–6.44 ± 0.09	–1.51 ± 0.03	1.23 ± 0.02	–1.54 ± 0.07	–1.42 ± 0.02
LA32/VA38/FL50	–5.06 ± 0.06	–0.68 ± 0.02	1.10 ± 0.01	–2.35 ± 0.06	–1.91 ± 0.01
LV32/VA38/FL50	–5.81 ± 0.04	–1.05 ± 0.01	1.19 ± 0.01	–1.91 ± 0.05	–1.69 ± 0.01
VT34	–7.78 ± 0.03	–3.07 ± 0.01	1.18 ± 0.01	–0.75 ± 0.04	–0.50 ± 0.01
VA34	–8.21 ± 0.03	–3.35 ± 0.07	1.22 ± 0.03	–0.49 ± 0.05	–0.33 ± 0.04
VG34	–6.05 ± 0.04	–1.67 ± 0.02	1.09 ± 0.01	–1.77 ± 0.05	–1.32 ± 0.01
TV36	–7.83 ± 0.08	–3.11 ± 0.03	1.18 ± 0.01	–0.71 ± 0.06	–0.47 ± 0.02
IA37Δ38	–5.25 ± 0.05	–0.18 ± 0.01	1.27 ± 0.01	–2.24 ± 0.05	–2.20 ± 0.01
VA38	–7.22 ± 0.03	–2.43 ± 0.01	1.20 ± 0.01	–1.08 ± 0.04	–0.87 ± 0.01
VA38/FL50	–6.31 ± 0.05	–1.59 ± 0.02	1.18 ± 0.01	–1.62 ± 0.05	–1.37 ± 0.01
VA38/FL50	–4.12 ± 0.06	0.32 ± 0.02	1.11 ± 0.01	–2.91 ± 0.05	–2.50 ± 0.01
TA39[c]	–8.39 ± 0.08	–3.13 ± 0.03	1.32 ± 0.01	–0.38 ± 0.06	0.46 ± 0.02
TD39[c]	–9.30 ± 0.17	–3.97 ± 0.07	1.35 ± 0.03	0.15 ± 0.11	0.04 ± 0.04

continued

Table 2. *continued*

Mutant	$\ln k_u^{H_2O}$	$\ln k_u^{4M}$	m_{ku} (M^{-1})	$\Delta\Delta G_{‡\text{-}F}^{H_2O}$[a] (kcal mol^{-1})	$\Delta\Delta G_{‡\text{-}F}^{4M}$[a] (kcal mol^{-1})
TA39 / EA41[c]	−7.99 ± 0.05	−2.82 ± 0.03	1.30 ± 0.01	−0.62 ± 0.05	0.64 ± 0.02
TD39 / EA41[c]	−8.95 ± 0.09	−3.65 ± 0.03	1.34 ± 0.02	−0.05 ± 0.07	0.15 ± 0.02
EA41[c]	−8.10 ± 0.08	−3.01 ± 0.03	1.28 ± 0.01	−0.56 ± 0.06	0.53 ± 0.02
RA43	−8.25 ± 0.11	−3.05 ± 0.04	1.30 ± 0.02	−0.47 ± 0.08	−0.50 ± 0.03
RA43 / DA45	−6.37 ± 0.23	−1.69 ± 0.08	1.17 ± 0.04	−1.58 ± 0.14	−1.31 ± 0.05
DA45	−8.08 ± 0.08	−3.01 ± 0.03	1.27 ± 0.01	−0.57 ± 0.06	−0.53 ± 0.02
VA47[b]	−3.08 ± 0.08	2.54 ± 0.03	1.41 ± 0.02	−3.53 ± 0.06	−3.81 ± 0.02
LA49[b]	−6.05 ± 0.15	−1.46 ± 0.05	1.15 ± 0.03	−1.77 ± 0.10	−1.45 ± 0.03
FL50	−7.13 ± 0.04	−2.03 ± 0.01	1.28 ± 0.01	−1.13 ± 0.05	−1.11 ± 0.01
FV50	−6.26 ± 0.04	−1.50 ± 0.01	1.19 ± 0.01	−1.64 ± 0.05	−1.42 ± 0.01
FA50	−5.12 ± 0.04	−0.33 ± 0.01	1.20 ± 0.01	−2.32 ± 0.05	−2.12 ± 0.01
VA51[b]	−7.20 ± 0.05	−2.21 ± 0.03	1.25 ± 0.01	−1.09 ± 0.05	−1.00 ± 0.02
DA52	−4.14 ± 0.07	−0.31 ± 0.01	0.96 ± 0.02	−2.90 ± 0.06	−2.13 ± 0.01
DN52	−8.93 ± 0.11	−3.89 ± 0.04	1.26 ± 0.02	−0.07 ± 0.08	−0.01 ± 0.02
ND56	−7.92 ± 0.07	−3.17 ± 0.02	1.19 ± 0.01	−0.66 ± 0.06	−0.43 ± 0.02
NA56	−8.12 ± 0.08	−3.00 ± 0.03	1.28 ± 0.01	−0.55 ± 0.06	−0.53 ± 0.02
IV57[b]	−9.31 ± 0.15	−3.56 ± 0.05	1.28 ± 0.03	0.16 ± 0.10	−0.20 ± 0.03
IA57[b]	−2.67 ± 0.16	1.94 ± 0.03	1.17 ± 0.04	−3.77 ± 0.10	−3.46 ± 0.02
AG58	−6.46 ± 0.06	−1.78 ± 0.02	1.17 ± 0.01	−1.52 ± 0.05	−1.26 ± 0.01
VT60	−8.42 ± 0.04	−3.44 ± 0.01	1.24 ± 0.01	−0.37 ± 0.05	−0.27 ± 0.01
VA60	−6.93 ± 0.08	−1.67 ± 0.03	1.32 ± 0.01	−1.25 ± 0.06	−1.32 ± 0.02
VG60	−3.93 ± 0.17	1.45 ± 0.02	1.34 ± 0.04	−3.02 ± 0.11	−3.17 ± 0.02
VM60 / ML39	−7.43 ± 0.04	−2.10 ± 0.01	1.33 ± 0.01	−0.95 ± 0.05	−1.07 ± 0.01
PA61	−4.03 ± 0.36	1.71 ± 0.04	1.43 ± 0.10	−2.97 ± 0.22	−3.32 ± 0.03
VT62	−7.35 ± 0.07	−2.28 ± 0.02	1.27 ± 0.01	−1.00 ± 0.06	−0.96 ± 0.02
VA82	−7.19 ± 0.06	−1.77 ± 0.02	1.36 ± 0.01	−1.09 ± 0.05	−1.26 ± 0.01
VG82	−3.98 ± 0.06	1.59 ± 0.01	1.39 ± 0.01	−3.00 ± 0.05	−3.25 ± 0.01

[a] $\Delta\Delta G_{‡\text{-}F} = -RT \ln(k_u / k'_u)$, where k_u and k'_u are the rate constants of unfolding for wild-type and mutant, respectively (in s^{-1}).
[b] From Jackson *et al*. (1993b).
[c] From Jackson & Fersht (1994).

state between wild-type and mutant, and k_f and k'_f are the rate constants of folding for the wild-type and mutant, respectively. The values of $\Delta\Delta G_{‡\text{-}U}$ are given in Table 3.

Analysis of two-state behaviour

The complete kinetics of folding and unfolding can be fitted to the equation (10), which is derived from equations (6) and (8), and based on a two-state transition:

$$\ln k = \ln(k_f^{H_2O} \exp(-m_{U\text{-}‡}[D]) + k_u^{H_2O} \exp(m_{‡\text{-}F}[D])) \quad (10)$$

For wild-type and the mutants, the kinetic data for unfolding and refolding fit well to this model (see Jackson *et al*., 1993b for representative curves). The values $k_f^{H_2O}$, $k_u^{H_2O}$, $m_{U\text{-}‡}$ and $m_{‡\text{-}F}$ obtained from this equation can be used to calculate equilibrium parameters, and can then be compared with those obtained directly from equilibrium measurements (Table 1). $\Delta G_{U\text{-}F}^{H_2O}$ can be calculated from the kinetic data using the ratios of the unfolding and refolding rate constants (Jackson *et al*., 1993b). $[D]_{50\%}$ can be calculated from kinetic experiments, using the equation:

$$[D]_{50\%} = (\ln(k_f^{H_2O} / k_u^{H_2O}) / (m_{ku} - m_{kf}) \quad (11)$$

We have determined m_{kf} between 0 and 0.6 M GdmCl, and also from 0.5 M GdmCl through the transition region. There are, on average, slightly higher values of m_{kf} for the range 0 to 0.6 M than calculated from the linear regions of the plots of ln k_f *versus* [GdmCl] for the higher concentrations, indicative of slight curvature in the plots of ln k_f *versus* [GdmCl]. Slight curvature in plots of $\Delta\Delta G_{U\text{-}F}$ *versus* [GdmCl] for protein denaturation has been noted by Santoro & Bolen (1992) at low [GdmCl] (<1.5 M), attributable to changes in ionic strength, but with excellent linearity at higher concentrations. In accord with this, we find that the values of m_{kf} at the higher concentrations fit the theoretical equations better, For example, a plot of $[D]_{50\%}$ calculated from equation (11) *versus* $[D]_{50\%}$ measured from equilibrium unfolding (Figure 1) has intercept −0.06(±0.04) M and slope 1.05(±0.01), using m_{kf} determined for [GdmCl] > 0.5 M, whereas just using data for m_{kf} in the range 0 to 0.6 M gives values of 0.27(±0.05) M and 1.01(±0.01), respectively. The values of $m_{U\text{-}F}$ for equilibrium denaturation (equation (1)) are related to those from kinetics by: $m_{U\text{-}F} = RT(m_{ku} - m_{kf})$. Substituting values of $m_{ku} - m_{kf}$ determined above 0.5 M GdmCl gives a mean value of 1.86(±0.01) kcal mol^{-1} M^{-1}, compared with the measured value of 1.90(±0.03). Using m_{kf} in the range 0 to 0.6 M GdmCl gives: $m_{U\text{-}F(calc)}$ = 2.05(±0.02).

There is a virtually perfect fit to a two-state mechanism for [GdmCl]〉 0.5 M, and an acceptable fit over the whole range. The slight positive

Table 3. Rates and thermodynamic data from refolding kinetics of CI2 mutants

Mutant	$\ln k_f^{H_2O}$	m_{ku}[a] (M^{-1})	$\Delta\Delta G_{\ddagger-U}^{H_2O}$[b] (kcal mol^{-1})
Wild-type	4.03 ± 0.04	−1.82 ± 0.12	0
KA2	4.21 ± 0.02	−2.24 ± 0.05	−0.11 ± 0.03
KM2	4.01 ± 0.01	−1.90 ± 0.05	0.02 ± 0.03
TA3	3.84 ± 0.01	−1.86 ± 0.03	0.11 ± 0.03
TV3	3.75 ± 0.02	−1.69 ± 0.06	0.17 ± 0.03
TG3	3.94 ± 0.02	−2.16 ± 0.06	0.06 ± 0.03
PA6	3.86 ± 0.01	−2.06 ± 0.04	0.10 ± 0.03
PA6/AG16	1.95 ± 0.04	−1.63 ± 0.11	1.23 ± 0.03
EA7	3.72 ± 0.01	−1.95 ± 0.02	0.19 ± 0.03
LA8[b]	3.36 ± 0.04	−2.02 ± 0.10	0.40 ± 0.03
KA11	3.59 ± 0.01	−1.62 ± 0.02	0.26 ± 0.03
SG12	3.65 ± 0.04	−2.33 ± 0.12	0.23 ± 0.04
SA12	3.39 ± 0.05	−2.05 ± 0.14	0.38 ± 0.04
EQ14	3.43 ± 0.03	−2.06 ± 0.08	0.36 ± 0.03
ED14	3.86 ± 0.04	−2.42 ± 0.11	0.10 ± 0.04
EN14	3.14 ± 0.02	−2.28 ± 0.05	0.53 ± 0.03
EQ15	3.62 ± 0.09	−2.23 ± 0.24	0.25 ± 0.06
ED15	3.76 ± 0.03	−2.17 ± 0.09	0.16 ± 0.03
EN15	3.07 ± 0.07	−2.13 ± 0.19	0.57 ± 0.05
EA14/EA15	3.10 ± 0.04	−1.69 ± 0.11	0.55 ± 0.04
SG12/EA14/EA15	2.73 ± 0.02	−1.84 ± 0.07	0.77 ± 0.03
SA12/EA14/EA15	2.90 ± 0.05	−2.14 ± 0.14	0.67 ± 0.04
AG16	2.09 ± 0.02	−2.46 ± 0.06	1.15 ± 0.03
KA17	3.81 ± 0.04	−2.20 ± 0.11	0.14 ± 0.04
KG17	2.56 ± 0.03	−2.04 ± 0.08	0.87 ± 0.03
KA18	3.90 ± 0.01	−1.77 ± 0.03	0.08 ± 0.03
KG18	2.88 ± 0.01	−1.86 ± 0.27	0.68 ± 0.03
VA19[c]	4.25 ± 0.02	−1.88 ± 0.04	−0.13 ± 0.03
IV20[c]	3.16 ± 0.03	−2.25 ± 0.07	0.52 ± 0.03
LA21	3.48 ± 0.02	−2.20 ± 0.06	0.33 ± 0.03
LG21	3.23 ± 0.03	−2.32 ± 0.08	0.48 ± 0.03
QA22	4.24 ± 0.02	−2.15 ± 0.05	−0.12 ± 0.03
QG22	3.92 ± 0.02	−2.27 ± 0.05	0.07 ± 0.03
DA23	4.43 ± 0.02	−2.16 ± 0.06	−0.23 ± 0.03
KA24	4.42 ± 0.05	−2.31 ± 0.13	−0.23 ± 0.04
KG24	3.52 ± 0.05	−2.47 ± 0.31	0.31 ± 0.04
PA25	3.44 ± 0.03	−2.50 ± 0.08	0.35 ± 0.03
EA26	3.81 ± 0.04	−2.20 ± 0.11	0.14 ± 0.04
IV29[c]	3.72 ± 0.03	−2.32 ± 0.07	0.19 ± 0.03
IA29[c]	2.37 ± 0.05	−2.73 ± 0.13	0.98 ± 0.04
IA29/IV57[c]	2.03 ± 0.05	−2.85 ± 0.13	1.18 ± 0.04
IV30	3.92 ± 0.02	−2.23 ± 0.06	0.07 ± 0.03
IA30	2.94 ± 0.05	−2.90 ± 0.14	0.65 ± 0.04
IG30	2.48 ± 0.06	−2.88 ± 0.16	0.92 ± 0.04
IT30	3.25 ± 0.05	−2.42 ± 0.18	0.47 ± 0.04
LA32	3.29 ± 0.11	−3.01 ± 0.37	0.44 ± 0.07
LI32	4.17 ± 0.02	−2.14 ± 0.06	−0.08 ± 0.03
LV32	4.07 ± 0.02	−2.15 ± 0.04	−0.02 ± 0.03
LV32/FL50	3.18 ± 0.04	−2.36 ± 0.10	0.51 ± 0.03
LV32/FA50	2.48 ± 0.03	−2.55 ± 0.10	0.92 ± 0.03
LA32/FL50	2.52 ± 0.04	−2.40 ± 0.12	0.90 ± 0.04
LA32/FA50	1.84 ± 0.11	−2.84 ± 0.28	1.30 ± 0.07
LA32/VA38	3.13 ± 0.10	−2.82 ± 0.28	0.54 ± 0.06
LV32/VA38	3.82 ± 0.04	−2.43 ± 0.10	0.13 ± 0.03
LA32/VA38/FL50	2.47 ± 0.07	−2.40 ± 0.19	0.92 ± 0.05
LV32/VA38/FL50	3.01 ± 0.07	−2.25 ± 0.18	0.60 ± 0.05
VT34	3.63 ± 0.02	−2.21 ± 0.06	0.24 ± 0.03
VA34	4.05 ± 0.03	−2.48 ± 0.10	−0.01 ± 0.03
VG34	3.36 ± 0.04	−2.48 ± 0.10	0.40 ± 0.03
TV36	3.79 ± 0.02	−2.29 ± 0.05	0.14 ± 0.03
IA37Δ38	3.69 ± 0.04	−2.65 ± 0.11	0.20 ± 0.03
VA38	3.73 ± 0.03	−2.40 ± 0.09	0.18 ± 0.03
VA38/FL50	2.90 ± 0.03	−2.58 ± 0.07	0.67 ± 0.03
VA38/FA50	2.02 ± 0.05	−2.76 ± 0.13	1.19 ± 0.04
TA39[d]	3.97 ± 0.03	−2.15 ± 0.08	0.04 ± 0.03
RA43	3.95 ± 0.03	−2.12 ± 0.08	0.05 ± 0.03
RA43/DA45	3.90 ± 0.05	−2.55 ± 0.14	0.08 ± 0.04
DA45	3.51 ± 0.02	−2.28 ± 0.06	0.31 ± 0.03
VA47[c]	2.31 ± 0.05	−2.56 ± 0.13	1.02 ± 0.04

continued

Table 3. *Continued*

Mutant	$\ln k_f^{H_2O}$	m_{ku}[a] (M^{-1})	$\Delta\Delta G_{\ddagger-U}^{H_2O}$[b] (kcal mol^{-1})
LA49[c]	0.61 ± 0.05	−2.81 ± 0.17	2.03 ± 0.04
FL50	3.03 ± 0.04	−2.70 ± 0.11	0.59 ± 0.04
FV50	3.03 ± 0.03	−2.31 ± 0.08	0.60 ± 0.03
FA50	2.08 ± 0.11	−2.73 ± 0.28	1.16 ± 0.07
VA51[c]	3.20 ± 0.04	−2.54 ± 0.11	0.49 ± 0.03
DA52	3.34 ± 0.38	−2.75 ± 0.11	0.41 ± 0.03
ND56	3.65 ± 0.04	−2.56 ± 0.10	0.23 ± 0.03
NA56	3.91 ± 0.02	−2.39 ± 0.06	0.07 ± 0.03
IV57[c]	3.85 ± 0.01	−2.01 ± 0.04	0.11 ± 0.03
IA57[c]	3.43 ± 0.02	−2.45 ± 0.07	0.36 ± 0.03
AG58	3.68 ± 0.02	−1.88 ± 0.06	0.21 ± 0.03
VT60	3.71 ± 0.03	−1.89 ± 0.09	0.19 ± 0.03
VA60	4.11 ± 0.04	−2.31 ± 0.11	−0.04 ± 0.04
VG60	3.82 ± 0.03	−2.10 ± 0.07	0.13 ± 0.03
VM60 / ML39	3.94 ± 0.05	−2.07 ± 0.13	0.05 ± 0.04
VT63	3.89 ± 0.01	−1.99 ± 0.01	0.09 ± 0.03
VA63	3.96 ± 0.02	−2.14 ± 0.05	0.04 ± 0.03
VG63	3.86 ± 0.04	−2.25 ± 0.11	0.10 ± 0.03

[a] Measured in the range 0 to 0.6 M GdmCl.
[b] $\Delta\Delta G_{\ddagger-U}^{H_2O} = -RT \ln(k_F / k_F')$, where k_f and k_f' are the rate constants of folding in water for wild-type and mutant, respectively, measured from pH-jump experiments (in s^{-1}).
[c] From Jackson *et al*. (1993b).
[d] From Jackson & Fersht (1994).

deviation in m_{kf} at lower [GdmCl] is not due to the two-state mechanism becoming three state, since the presence of an intermediate leads to a lowering of m_{kf}, as illustrated for barnase by Matouschek *et al*. (1990). Instead, the deviation is likely to result from a genuine non-linearity of response of free energy to [GdmCl], as demonstrated by Santoro & Bolen (1992). The value of m_{kf} is not used in any of the calculations used for analysing the transition state in this study; the only experimental quantities used are: $k_f^{H_2O}$, which is measured directly in water; k_u^{4M}, which is obtained by a short extrapolation; and $\Delta\Delta G_{U-F}^{[D]_{50\%}}$, which has been checked by differential scanning calorimetry. In other studies, where we wish to see how m_{kf} changes on mutation, we prefer the data set determined for 0 to 0.6 M GdmCl, since it measured the more precisely, and systematic deviations tend to cancel out during comparisons.

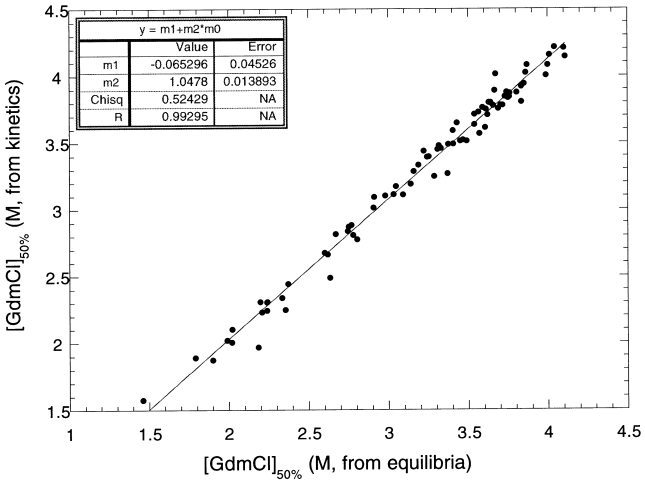

Figure 1. Plot of the concentration of GdmCl required for 50% denaturation of CI2 calculated from the ratio of kinetic rate constants (equation (11)) against that measured by equilibrium denaturation, showing the excellence of the fit to two-state kinetics.

The protein engineering method

The procedure we use for studying folding pathways by protein engineering has been discussed extensively elsewhere (Fersht *et al*., 1992; Fersht, 1993, 1995a). Briefly, a suitable side-chain interaction is truncated by mutagenesis and the resulting mutant characterized by equilibrium denaturation and kinetic unfolding and refolding experiments to determine the extent of this interaction at different stages on the folding pathway. The interaction is determined by the following ratio:

$$\Phi_F = \Delta\Delta G_{\ddagger-U} / \Delta\Delta G_{F-U} \qquad (12)$$

where $\Delta\Delta G_{\ddagger-U}$ is the change in free energy of folding of the transition state on mutation, and $\Delta\Delta G_{F-U}$ is the change in equilibrium free energy of folding on mutation. A Φ_F of 1 shows that the transition state is disrupted by mutation by the same energy as is the fully folded protein and so indicates complete formation of native structure in the transition state, while $\Phi_F = 0$ shows that the transition state is as insensitive to mutation as is the fully denatured state and so is indicative of complete lack of native structure in the transition state. Fractional Φ-values can be difficult to interpret for several reasons. They can arise if a protein folds by parallel pathways, in which parts of the protein are native-like in the transition-state of one pathway (i.e. $\Phi_F = 1$) but unfolded in the transition-state of another pathway (i.e. $\Phi_F = 0$). However, CI2 was shown recently to fold along a single pathway involving one or at the

most a closely related set of transition states (Fersht *et al*., 1994).

Analysis of fractional Φ-values can also be complicated by factors such as access of water to the site of mutation. This influences the observed Φ-value by introducing solvation energy terms during the folding reaction. In the case where water does not enter the site of mutation, there may be a more linear relationship between the Φ-value and the extent of formation of non-polar interactions (Fersht *et al*., 1992). Double mutant cycles allow us to focus on the specific interaction between two residues in which the effect of surrounding residues and solvent interactions tend to cancel out (Horovitz & Fersht, 1990; Fersht *et al*., 1992). Interactions of individual moieties within a side-chain can be probed during folding by making a series of mutations at a single site (Fersht *et al*., 1992; Serrano *et al*., 1992a). For example, the series of mutations Val → Ala → Gly allows us to attribute structure formation to each of the C^{β}, $C^{\gamma 1}$ and $C^{\gamma 2}$ methyl(ene) groups.

Choice of mutation

Our approach for designing mutations is as follows. The pseudo-wild-type crystal structure of CI2 (Harpaz *et al*., 1994) is examined for side-chain atoms that probe a particular structural interaction. The ideal mutation, which has been termed ''non-disruptive'' (Fersht *et al*., 1992), deletes only a small part of the side-chain, removing defined interactions without introducing new ones, altering the stereochemistry or perturbing the structure. Thus it acts as a probe of the folding pathway. Our favourite mutations are: Ile → Val → Ala →Gly; Thr → Ser; Ser → Ala; Tyr → Phe. When in doubt, we mutate larger side-chains to Ala.

Since Φ-value analysis monitors formation of interactions of side-chains, it is not a direct measure of the extent of secondary structure formation, but Φ-values can probe the consequences of formation of secondary structure. We have made, for example, a series of mutations at sites in the α-helix and in the β-sheet where the side-chain makes interactions almost exclusively within the secondary structural element, and thus the Φ-values which we obtain at these sites are an indirect probe of the formation of secondary structure interactions. In addition, we have made mutations at side-chains whose interactions are predominantly tertiary, for example within the core and the minicore. The results should, therefore, provide us with a very comprehensive picture of the transition state in the folding pathway of CI2.

Structure of CI2

The secondary structure of CI2, defined from an NMR analysis of the solution structure (Ludvigsen *et al*., 1991) and in agreement with the crystal structure of the mutant EA14EA15 (which we call the pseudo-wild-type, Harpaz *et al*., 1994) is as follows (Figure 2): residues 3 to 5, β-strand 1; 5 to 8, type III reverse turn; 8 to 11, type II reverse turn; residues 10 to 11, β-strand 2; residues 12 to 24, α-helix; residues 25 to 28, type I reverse turn; 28 to 34 β-strand 3; 35 to 44 reactive site loop (extended structure); residues 45 to 52, β-strand 4; residues 52 to 54, turn; residues 55 to 58, β-strand 5 and residues 60 to 64, β-strand 6. The numbering is based on the truncated version of CI2 lacking the first 19 residues found in the original CI2. These residues are devoid of fixed structure and their removal does not change the folding properties of CI2.

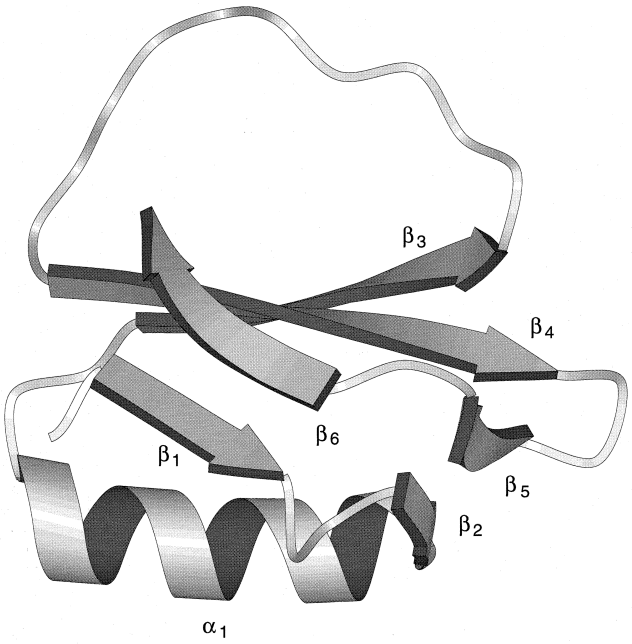

Figure 2. Schematic representation of the structure of CI2. The diagram was produced using the program MolScript (Kraulis, 1991).

The β-sheet packs against the α-helix to form the hydrophobic core. The reactive-site loop projects out from the other side of the β-sheet, in an extended conformation. In contrast to the hydrophobic character of the interface of the β-sheet with the α-helix, the other side of the sheet forms an extensive network of hydrogen bonds and electrostatic interactions with the reactive-site loop.

α-Helix

The α-helix runs from 13 to 23 according to Φ and Ψ angles (Li & Daggett, 1994), but we use the definition of caps by Richardson & Richardson (1988) to designate Ser12 as the N-cap of the helix and Lys24 as its C-cap. The α-helix consists of three turns and is somewhat irregular in its geometry. From the crystal structure (McPhalen & James, 1987) and from the recent higher-resolution crystal structure of the pseudo-wild-type (Harpaz *et al*., 1994), it appears that some of the main-chain NH–CO hydrogen bonds are longer than is usual for an α-helix, while others are not formed at all.

These are NH18–CO14, NH19–CO15, NH22–CO18 and NH23–CO19. The N-cap of the helix at position 12 is occupied by a Ser residue. The OH group of the side-chain of Ser12 makes a hydrogen bond with the main-chain NH group of residue 15 (N-cap + 3) and with a water molecule. This end of the helix is very solvent-exposed, with many strong hydrogen bonds between main-chain CO and NH groups and water molecules. On the buried side of the α-helix, there are four hydrophobic residues (Val13, Ala16, Val19, Ile20) which pack against the β-sheet to form the hydrophobic core. On the solvent-exposed face are residues Lys17, Lys18, Leu21 and Gln22. The C-cap at position 24 is occupied by a Lys residue, followed by an α-helix stop at Pro25. This end of the helix is much less solvated than the other, with the C-cap almost completely buried.

β-Sheet

The six-stranded β-sheet of CI2 is classified as a pseudo-β-sheet because the components of the main pair of strands (strands 3 and 4) are joined by a β-bridge (the two β-sheet hydrogen bonds Ala58–Phe50 and Gly64–Arg46) to the other strands. The two parallel strands 3 and 4 (28 to 34 and 45 to 51) form the main body of the β-sheet, displaying a regular β-strand hydrogen bond pattern. Shorter segments of anti-parallel structures are formed between 55-57 in strand 5 and 11-13 in strand 2, 56-58 in strand 5 and 50-52 in strand 4, 62-64 in strand 6 and 46-48 in strand 4 as well as between 61-63 in strand 6 and 3-5 in strand 1. The β-sheet shows deviation from canonical structure in that an array of ordered water molecules are observed in the crystal structure which mediate hydrogen bonding between strand 4 and strand 6. These bridging water molecules are not visible in the solution structure determined by NMR (Ludvigsen *et al*., 1991), where the backbone of Arg48 forms a hydrogen bond with that of Gly64 directly, rather than by bridging water molecules seen in the crystal structure. A recent molecular dynamics study (Nanzer *et al*., 1994) demonstrates the dynamic nature of the β-sheet structure of CI2: the hydrogen bond between the N-H of residue 48 and the C = O of residue 64 is present only some 20% of the time. The hydrogen bonding between two strands comprising residues 28 to 30 (strand 3) and 45 to 47 (strand 4) as well as residues 3 to 5 (strand 1) and 61 to 63 (strand 6) was also shown to be mobile: part of the time they form a four-stranded β-sheet in the core, and part of the time the strands move apart to form hydrogen bonds with nearby water molecules (Nanzer *et al*., 1994).

Hydrophobic cores

The hydrophobic core is formed by packing of one face of the α-helix onto the β-sheet, and consists of 12 hydrophobic residues, namely Trp5, Leu8, Ala16, Val19, Ile20, Ala27, Ile29, Val47, Leu49, Val51, Ile57 and Pro61. The core buries almost 1900 Å^2 of surface, and exposes only 133 Å^2 to solvent. In addition, the side-chains of residues Leu32, Val38 and Phe50 form a hydrophobic pocket near one end of the reactive loop, on the other side of the β-sheet relative to the hydrophobic core. This cluster is more solvent-accessible than the main core, burying 425 Å^2 and exposing to solvent 133 Å^2 of hydrophobic surface area. We call this cluster the minicore. These three residues are highly conserved among different members of the inhibitor family to which CI2 belongs.

Turns and ends

Hydrogen bonds are prominent in the turns and in the terminal residues of CI2. While the free amino group of Met1, an artifact of the truncated construct which we are studying, does not link to any group, the N^{ζ} atom of Lys2 is bonded with $O^{\epsilon 2}$ of Glu7, and this interaction (or the intra-residual hydrogen bond between $O^{\epsilon 1}$ and N of Glu7) may contribute to the stability of the turn after β-strand 1. The pK_a of Glu7 is lowered from 4.3 to 3.1 (B. Davis & A.R.F., unpublished). In the reactive site loop, the backbone of Val34 and Gly35 forms a small bulge out of the loop, and this appears to be stabilized by a hydrogen bond between $O^{\gamma 1}$ of Thr36 and the carbonyl oxygen of Pro33. At the other end of the loop, Arg43 and Asp45 form two hydrogen bonds between their respective side-chains. The charged carboxyl group of the C-terminal Gly64 forms a strong hydrogen bond with N^{H2} of Arg46. In addition, N^{ϵ} of Arg46 also hydrogen bonds with the O of Gly64. Finally, there is a very extensive hydrogen-bonding network in the β-hairpin loop joining β-strands 4 and 5, where the side-chain $O^{\delta 1}$ of Asp52 forms simultaneous hydrogen bonds with the amides of Leu54 and Asn56 (as well as Asp55 in the wild-type crystal structure). This network decreases the pK_a of the side-chain from 3.9 to 2.7 (B. Davis & A.R.F., unpublished).

Structure of the transition state

Relationship of Φ_F and Φ_U

The two-state nature of the folding kinetics of CI2 means that the strength of interactions in the transition state can be studied by kinetics directly from both directions of folding (Φ_F from refolding kinetics in water) and unfolding (Φ_U by extrapolating the unfolding kinetics to zero molar denaturant (water), where $\Phi_U = \Delta\Delta G_{\ddagger\text{-}F} / \Delta\Delta G_{U\text{-}F}$), and thus provides us with independent means of verifying results and a means of checking some of the assumptions made in the analysis. For a two-state reaction, $\Phi_F + \Phi_U = 1$, so that $\Phi_F = 1 - \Phi_U$.

The Φ-values obtained from these two approaches, Φ_F and Φ_U, respectively, are listed in Table 4. $\Phi_F^{H_2O}$ is measured accurately with no extrapolation, and so is the most reliable quantity. $1 - \Phi_U^{H_2O}$ is inaccurate due to a long extrapolation

Table 4. Values of Φ_F and Φ_U calculated from refolding and unfolding kinetics, respectively

Mutant	$\Phi_F^{H_2O}$[a]	$1-\Phi_U^{H_2O}$[b] (=$\Phi_F^{H_2O}$)	$1-\Phi_U^{4M}$[c] (=Φ_F^{4M})
Core			
LA8[d]	0.15 ± 0.01	–0.02 ± 0.04	0.24 ± 0.05
AG16	1.06 ± 0.05	1.24 ± 0.08	1.20 ± 0.03
VA19[d]	–0.26 ± 0.07	–0.51 ± 0.26	–0.71 ± 0.14
IV20[d]	0.40 ± 0.03	0.54 ± 0.06	0.56 ± 0.04
IV29[d]	0.17 ± 0.03	0.44 ± 0.06	0.39 ± 0.04
IA29[d]	0.25 ± 0.01	0.28 ± 0.03	0.43 ± 0.05
IA29/IV57[d]	0.29 ± 0.01	0.32 ± 0.03	0.39 ± 0.03
VA47[d]	0.21 ± 0.01	0.28 ± 0.03	0.14 ± 0.11
LA49[d]	0.53 ± 0.02	0.54 ± 0.03	0.63 ± 0.02
VA51[d]	0.25 ± 0.02	0.45 ± 0.03	0.54 ± 0.04
IV57[e]	—	—	—
IA57[d]	0.08 ± 0.01	0.12 ± 0.03	0.18 ± 0.09
PA61	0.02 ± 0.01	0.11 ± 0.07	–0.07 ± 0.08
Minicore			
LA32	0.19 ± 0.03	0.33 ± 0.06	0.41 ± 0.01
LI32	–0.31 ± 0.14	–0.16 ± 0.36	–0.55 ± 0.09
LV32	–0.04 ± 0.05	0.04 ± 0.15	–0.22 ± 0.07
LV32/FL50	0.21 ± 0.01	0.30 ± 0.03	0.46 ± 0.03
LV32/FA50	0.27 ± 0.01	0.31 ± 0.03	0.41 ± 0.03
LA32/FL50	0.26 ± 0.01	0.31 ± 0.02	0.42 ± 0.03
LA32/FA50	0.27 ± 0.02	0.33 ± 0.03	0.33 ± 0.07
LA32/VA38	0.17 ± 0.02	0.31 ± 0.02	0.37 ± 0.02
LV32/VA38	0.07 ± 0.02	0.17 ± 0.04	0.20 ± 0.03
LA32/VA38/FL50	0.26 ± 0.01	0.32 ± 0.02	0.45 ± 0.02
LV32/VA38/FL50	0.22 ± 0.02	0.30 ± 0.02	0.34 ± 0.02
VA38	0.12 ± 0.02	0.27 ± 0.04	0.46 ± 0.03
VA38/FL50	0.26 ± 0.01	0.37 ± 0.02	0.49 ± 0.02
VA38/FL50	0.26 ± 0.01	0.33 ± 0.02	0.49 ± 0.03
FL50	0.28 ± 0.02	0.47 ± 0.03	0.48 ± 0.03
FV50	0.25 ± 0.01	0.31 ± 0.03	0.48 ± 0.04
FA50	0.30 ± 0.02	0.40 ± 0.02	0.51 ± 0.03
Helix			
SG12	0.29 ± 0.05	0.54 ± 0.18	0.54 ± 0.06
SA12	0.43 ± 0.05	0.66 ± 0.16	0.63 ± 0.06
EQ14	1.23 ± 0.27	0.86 ± 0.29	0.57 ± 0.09
ED14	0.20 ± 0.07	0.69 ± 0.34	0.43 ± 0.11
EN14	0.75 ± 0.06	0.97 ± 0.13	0.68 ± 0.05
EQ15	0.53 ± 0.14	0.58 ± 0.22	0.50 ± 0.07
ED15	0.22 ± 0.05	0.39 ± 0.10	0.23 ± 0.06
EN15	0.53 ± 0.05	0.39 ± 0.07	0.50 ± 0.04
EA14/EA15	0.73 ± 0.08	0.32 ± 0.20	0.32 ± 0.08
SG12/EA14/EA15	0.47 ± 0.03	0.14 ± 0.11	0.36 ± 0.06
SA12/EA14/EA15	0.40 ± 0.02	0.29 ± 0.05	0.46 ± 0.04
KA17	0.28 ± 0.08	–0.48 ± 0.19	–0.42 ± 0.07
KG17	0.38 ± 0.02	0.26 ± 0.04	0.44 ± 0.03
KA18[e]	—	—	—
KG18	0.70 ± 0.06	0.58 ± 0.09	0.48 ± 0.05
LA21	0.25 ± 0.03	0.28 ± 0.06	0.49 ± 0.04
LG21	0.35 ± 0.03	0.21 ± 0.05	0.36 ± 0.03
QA22[e]	—	—	—
QG22	0.12 ± 0.05	0.25 ± 0.13	0.27 ± 0.06
DA23	–0.25 ± 0.03	0.28 ± 0.14	0.01 ± 0.05
KA24	–0.35 ± 0.07	–0.04 ± 0.13	–0.34 ± 0.07
KG24	0.10 ± 0.01	0.16 ± 0.04	–0.01 ± 0.10
β-Sheet			
TA3	0.13 ± 0.03	0.26 ± 0.10	0.13 ± 0.05
TV3	0.52 ± 0.14	0.17 ± 0.24	–0.05 ± 0.06
TG3	0.05 ± 0.03	0.30 ± 0.10	0.29 ± 0.05
IV30[e]	—	—	—
IA30	0.31 ± 0.02	0.57 ± 0.03	0.66 ± 0.02
IG30	0.26 ± 0.01	0.37 ± 0.03	0.43 ± 0.02
IT30	0.35 ± 0.03	0.63 ± 0.03	0.72 ± 0.02
VT34	0.23 ± 0.03	0.27 ± 0.06	0.44 ± 0.03
VA34	–0.01 ± 0.05	0.23 ± 0.15	0.42 ± 0.09
VG34	0.16 ± 0.01	0.27 ± 0.03	0.48 ± 0.02
AG58	0.11 ± 0.02	0.19 ± 0.04	0.41 ± 0.06
VT60	0.51 ± 0.17	0.03 ± 0.31	0.14 ± 0.08
VA60	–0.03 ± 0.02	0.18 ± 0.05	0.12 ± 0.05

continued

Table 4. *continued*

Mutant	$\Phi_F^{H_2O}$[a]	$1-\Phi_U^{H_2O}$[b] ($=\Phi_F^{H_2O}$)	$1-\Phi_U^{4M}$[c] ($=\Phi_F^{4M}$)
VG60	0.04 ± 0.01	0.06 ± 0.04	0.27 ± 0.03
VM60 / ML39	0.05 ± 0.03	0.16 ± 0.05	0.08 ± 0.04
VT63	0.08 ± 0.02	0.13 ± 0.07	0.09 ± 0.07
VA63	0.03 ± 0.02	0.25 ± 0.05	0.13 ± 0.06
VG63	0.03 ± 0.01	0.14 ± 0.02	0.14 ± 0.03
Reactive site loop			
TV36	0.19 ± 0.04	0.06 ± 0.13	0.29 ± 0.06
IA37 / Δ38	0.09 ± 0.01	0.22 ± 0.02	0.22 ± 0.02
TA39	0.19 ± 0.07	0.46 ± 0.11	0.33 ± 0.01
EA41	0.32 ± 0.09	0.20 ± 0.15	0.15 ± 0.16
TA39 / EA41	0.18 ± 0.06	0.30 ± 0.07	0.27 ± 0.06
TD39 / EA41	0.82 ± 0.24	0.80 ± 0.25	0.38 ± 0.17
YG42	0.07 ± 0.01	0.16 ± 0.04	0.08 ± 0.03
RA43	0.09 ± 0.06	0.19 ± 0.17	0.05 ± 0.07
RA43 / DA45	0.06 ± 0.03	–0.29 ± 0.14	–0.36 ± 0.11
DA64	0.44 ± 0.03	0.29 ± 0.09	0.30 ± 0.04
Turns			
KA2	–0.19 ± 0.05	–0.30 ± 0.14	–0.39 ± 0.05
KA2 / EA7	0.17 ± 0.02	0.10 ± 0.13	0.14 ± 0.05
KM2	0.03 ± 0.04	0.29 ± 0.14	0.11 ± 0.05
PA6	0.07 ± 0.02	0.47 ± 0.19	0.44 ± 0.12
EA7	0.40 ± 0.07	0.44 ± 0.19	0.30 ± 0.07
KA11	–0.49 ± 0.07	0.51 ± 0.13	0.26 ± 0.06
PA25	0.20 ± 0.02	0.12 ± 0.12	0.33 ± 0.04
EA26	0.42 ± 0.12	0.22 ± 0.22	0.18 ± 0.07
DA52	0.12 ± 0.01	0.15 ± 0.03	0.39 ± 0.06
ND56	0.19 ± 0.03	0.45 ± 0.05	0.61 ± 0.02
NA56	0.09 ± 0.04	0.34 ± 0.08	0.30 ± 0.03

Φ_F and Φ_U are related by the expression $\Phi_F + \Phi_U = 1$, so that $\Phi_F = 1 - \Phi_U$. $\Phi_F^{H_2O}$ from the folding experiments is the most accurate, since data measured in H_2O are directly compared. Φ_U^{4M} is less accurate, since unfolding data are extrapolated to 4 M GdmCl. $\Phi_U^{H_2O}$ is the least accurate, since there is a long extrapolation of unfolding data at high [GdmCl] to 0 M. The presence of 4 M GdmCl is expected to perturb the structure of the transition state.

[a] $\Phi_F^{H_2O} = \Delta\Delta G_{\ddagger\text{-}U}^{H_2O} / \Delta\Delta G_{F\text{-}U}^{[D]_{50\%}}$ (where $\Delta\Delta G_{F\text{-}U}^{[D]_{50\%}} = -\Delta\Delta G_{U\text{-}F}^{[D]_{50\%}}$).

[b] $\Phi_U^{H_2O} = \Delta\Delta G_{\ddagger\text{-}F}^{H_2O} / \Delta\Delta G_{U\text{-}F}^{[D]_{50\%}}$.

[c] $\Phi_U^{4M} = \Delta\Delta G_F^{4M} / \Delta\Delta G_{U\text{-}F}^{4M}$.

[d] From Jackson & Fersht (1993b).

[e] Φ-Values were not calculated for these mutants because $\Delta\Delta G_{U\text{-}F}^{[D]_{50\%}} \approx 0$.

from the concentrations of GdmCl over which the rates of unfolding were measured. $1-\Phi_U^{4M}$ is accurate, being determined only from a small extrapolation from higher values of [GdmCl], but is measured in 4 M denaturant. It is expected that there will be some perturbation of the structure of the transition state in 4 M GdmCl because of the presence of denaturant that stabilizes the denatured state of proteins. In particular, there should be a general effect according to the Hammond postulate (Hammond, 1955), which states that as a state on a reaction pathway becomes destabilized, the transition state for its formation should move closer to it in structure. This postulate holds for protein folding (Matouschek & Fersht, 1993) and so it is likely that the structure of the transition state for protein folding moves closer to that of the folded protein with increasing concentration of denaturant, as has been found for barnase (Matthews & Fersht, 1995). The values of $\Phi_F^{H_2O}$ and $1-\Phi_U^{H_2O}$ should be identical for CI2 folding according to the two-state model, and there is, indeed, good agreement within experimental error for the majority of mutations, with the best agreement observed for mutations in the hydrophobic core and minicore. In general, the values of $1-\Phi_U^{4M}$ are somewhat higher than $\Phi_F^{H_2O}$, as expected. The differences are not sufficiently high to change radically the description of energies. Poorer agreement may result where there are smaller equilibrium destabilization energies in certain regions of the protein on mutation and the consequent large errors in the calculation of Φ-values, rather than deviation from the two-state behaviour.

The major assumptions of the protein engineering approach are that (1) the mutation does not alter the folding pathway, (2) the mutation does not significantly change the structure of the folded protein, (3) the mutation does not perturb the structure of the unfolded state, and (4) the target groups do not make new interactions with new partners during the course of the reaction (Fersht *et al*., 1992). Evidence that these assumptions are valid for CI2 is provided by the observation of good agreement between Φ-values when different mutations are constructed at the same site. Thus, the overall formation of structure in the transition state parallels the formation of interactions between individual side-chains. We now describe the structures of individual regions of CI2 in the

transition state, as elucidated from the protein engineering approach.

Hydrophobic core

Some of these results have been described previously (Jackson *et al*., 1993b). The results for the major hydrophobic core of CI2 can be split into three categories: those mutations that result in a Φ_F of close to zero, those mutations that result in fractional Φ_F, and one mutation that results in a Φ_F of 1. The residues with low Φ_F (between 0 and 0.2) are located on the edge of the core, in turns or at the beginning or end of a β-strand. These include the following mutations: Leu → Ala8, Val → Ala47, Ile → Ala57, Pro → Ala61, in β-strand 6. The residues with fractional values of Φ_F (between 0.3 and 0.55) are all located in the centre of the core, either in the α-helix or in the centre of a β-strand. These include the following mutations: Ile → Val20, Ile → Val29, Ile → Ala29, Leu → Ala49, Val → Ala51, Ile → Ala29/Ile → Val57. Val → Ala19 has a negative Φ_F-value, suggesting that the side-chain of Val19 makes more contacts, some of them non-native, in the transition state than in the native state. There must, therefore, be some structural rearrangement on going from the transition state to the native state. Ala16, in the second turn of the α-helix, is completely buried and probes interactions with the first turn of the α-helix and β-strand 4. The mutation Ala → Gly16 has a Φ_F of 1, a result which has not been found for any other residue of CI2, including the residues with which Ala16 interacts in the native state (Ile57, Leu8 and Leu49). However, while the mutation Ala → Gly16 very specifically probes the interactions involving just the methyl group (C^β), the mutations that have been made at the residues with which it interacts (Ile → Ala57, Leu → Ala8 and Leu → Ala49) delete several atoms and, thus, may obscure very specific atom-to-atom interactions during folding.

Minicore

All values of Φ_F for mutations in the minicore are fractional, and most lie between 0.2 and 0.4. Fractional Φ-values are difficult to interpret because they may arise from a variety of effects. However, the rate constants for the 16 minicore mutants and wild-type protein fit beautifully to a Brønsted plot of $\ln k_f$ (or $\ln k_u$) *versus* $\Delta G_{F\text{-}U}/RT$ (Fersht *et al*., 1994) with a Brønsted β-value ($=RT\partial \ln k_f/\partial \Delta G_{F\text{-}U}$) of 0.3. This uniform response of an element of structure to a large number of mutations is best explained by the structure having 30% of energy of interaction in the transition state (Fersht *et al*., 1994). This interpretation is supported by analysis using double mutant cycles (Carter *et al*., 1984; Horovitz & Fersht, 1990, 1992; Fersht *et al*., 1992). The coupling energy, $\Delta\Delta G_{int}$, between two residues can be calculated from a double-mutant cycle. (The coupling energy between two residues is the energy change on mutating two residues simultaneously minus the sum of the change in energies on mutating them separately, and is the quantitative measure of the co-operativity of the two residues concerned in stabilising a structure). The coupling energy between two residues in the transition state may also be calculated from kinetics and manipulated to give a Φ-value, Φ_F^{int}, (Table 7, Horovitz *et al*., 1991; Fersht *et al*., 1992; Horovitz & Fersht, 1992). Values of Φ_F^{int} also cluster around 0.3 showing that the co-operativity of formation over this region is about 30% of the energetics in the fully folded state.

The families of mutants constructed at single sites allow us to perform a fine-structure analysis, in which Φ_F for specific parts of a single side-chain in

Table 5. Coupling energies for double mutants in CI2

Triple mutant	$\Delta\Delta G_{int}^{U\text{-}F}$ ([D]50%) (kcal mol^{-1})	$\Delta\Delta G_{int}^{\ddagger\text{-}U}$ (4 M GdmCl) (kcal mol^{-1})	$\Delta\Delta G_{int}^{\ddagger\text{-}U}$ (H_2O) (kcal mol^{-1})	$\Phi_{int}^{F(H_2O)}$[a]	$\Phi_{int}^{F(4M)}$[a]
KA2/EA7	–0.08 ± 0.07	0.16 ± 0.05	–0.02 ± 0.05	[b]	[b]
KA2/DA23	0.35 ± 0.08	0.23 ± 0.05	–0.06 ± 0.05	–0.17 ± 0.15	0.35 ± 0.11
RA43/DA45	0.15 ± 0.11	–0.28 ± 0.06	0.28 ± 0.04	[b]	[b]
LV32/FL50	0.19 ± 0.10	0.24 ± 0.02	0.23 ± 0.07	[b]	[b]
LV32/FA50	0.93 ± 0.13	0.26 ± 0.03	0.30 ± 0.07	0.32 ± 0.08	0.28 ± 0.04
LA32/FL50	1.06 ± 0.11	0.32 ± 0.04	0.40 ± 0.07	0.38 ± 0.07	0.40 ± 0.05
LA32/FA50	1.42 ± 0.21	0.43 ± 0.04	0.55 ± 0.07	0.39 ± 0.07	0.30 ± 0.05
VA38/FL50	1.00 ± 0.10	0.39 ± 0.02	0.22 ± 0.07	0.22 ± 0.07	0.39 ± 0.04
VA38/FA50	0.94 ± 0.13	0.45 ± 0.02	0.32 ± 0.05	0.34 ± 0.08	0.48 ± 0.07
LA32/VA38	0.68 ± 0.10	0.30 ± 0.04	0.31 ± 0.06	0.45 ± 0.11	0.44 ± 0.08
LV32/VA38	0.13 ± 0.09	0.12 ± 0.02	0.19 ± 0.06	[b]	[b]

Derived from double mutant cycles. The cycle indicated by KA2/EA7, for example, consists of wild-type, the single mutants Lys → Ala2 and Glu → Ala7 and the double mutant Lys → Ala2/Glu → Ala7. The coupling energy, $\Delta\Delta G_{int}^{U\text{-}F}$, is calculated from: $\Delta\Delta G_{int}^{U\text{-}F} = \Delta\Delta G_{E-XY\rightarrow E-Y}^{U\text{-}F} + \Delta\Delta G_{E-XY\rightarrow E-X}^{U\text{-}F} - \Delta\Delta G_{E-XY\rightarrow E}^{U\text{-}F}$, where the subscript $E-XY \rightarrow E-Y$ indicates the mutant Lys → Ala21, $E-XY \rightarrow E-X$ indicates the mutant Glu → Ala26, etc. (Fersht *et al*., 1992). $\Delta\Delta G_{int}^{\ddagger\text{-}F}$ (H_2O) and $\Delta\Delta G_{int}^{\ddagger\text{-}F}$ (4 M GdmCl) are calculated from the values of $\Delta\Delta G^{\ddagger\text{-}U}$ in a similar manner (Fersht *et al*., 1992). $\Delta\Delta G_{int}^{\ddagger\text{-}U}$ at 4 M GdnHCl is calculated from: $\Delta\Delta G_{int}^{\ddagger\text{-}U} = \Delta\Delta G_{int}^{U\text{-}F} - \Delta\Delta_{int}^{\ddagger\text{-}F}$.

[a] $\Phi_{int}^{F} = \Delta\Delta G_{int}^{\ddagger\text{-}U}/\Delta\Delta G_{int}^{U\text{-}F}$.

[b] Φ-Values are not calculated for these mutants, because $\Delta\Delta G_{U\text{-}F}^{[D]50\%} \approx 0$.

Table 6. Fine-structure analysis of mutants in CI2

Mutation	$\Delta\Delta G_{U\ddagger F}^{H_2O}$ (kcal mol^{-1})	$\Delta\Delta G_{\ddagger F}^{4M}$ (kcal mol^{-1})	$\Delta\Delta G_{\ddagger U}^{H_2O}$ (kcal mol^{-1})	$1 - \Phi_U^{4M}$	$\Phi_F^{H_2O}$
VG3	0.83 ± 0.09	0.53 ± 0.04	–0.11 ± 0.04	0.36 ± 0.08	–0.14 ± 0.05
VA3	0.52 ± 0.09	0.40 ± 0.02	–0.06 ± 0.04	0.23 ± 0.13	–0.11 ± 0.08
VA29	2.80 ± 0.10	1.68 ± 0.03	0.80 ± 0.05	0.40 ± 0.02	0.29 ± 0.02
VA30	2.20 ± 0.10	0.68 ± 0.04	0.58 ± 0.05	0.69 ± 0.02	0.27 ± 0.03
AG30	1.40 ± 0.09	1.38 ± 0.03	0.27 ± 0.06	0.02 ± 0.07	0.19 ± 0.04
VT30	1.42 ± 0.09	0.26 ± 0.04	0.40 ± 0.05	0.82 ± 0.03	0.28 ± 0.04
IV32	0.25 ± 0.09	0.21 ± 0.02	0.06 ± 0.04	0.16 ± 0.32	0.23 ± 0.18
VA32	1.87 ± 0.08	0.96 ± 0.04	0.46 ± 0.08	0.49 ± 0.03	0.25 ± 0.04
VA3 / FL50	1.00 ± 0.09	0.37 ± 0.02	0.39 ± 0.05	0.62 ± 0.04	0.39 ± 0.06
VA32 / VA38	1.31 ± 0.08	0.58 ± 0.02	0.41 ± 0.07	0.56 ± 0.03	0.31 ± 0.06
VA32 / VA38 / FL50	0.76 ± 0.09	0.22 ± 0.02	0.32 ± 0.07	0.71 ± 0.04	0.42 ± 0.10
TA34	–0.39 ± 0.12	–0.17 ± 0.04	–0.25 ± 0.04	0.57 ± 0.17	0.63 ± 0.23
LV50	0.28 ± 0.09	0.31 ± 0.02	0.00 ± 0.05	–0.12 ± 0.38	0.01 ± 0.17
VA50	1.45 ± 0.11	0.69 ± 0.02	0.56 ± 0.07	0.52 ± 0.04	0.39 ± 0.06
LA50	1.73 ± 0.10	1.01 ± 0.02	0.57 ± 0.08	0.42 ± 0.04	0.33 ± 0.05
LA50 / VA38	1.79 ± 0.11	1.13 ± 0.02	0.41 ± 0.07	0.37 ± 0.04	0.23 ± 0.04
LA50 / LA32	1.37 ± 0.20	0.66 ± 0.02	0.40 ± 0.08	0.52 ± 0.07	0.29 ± 0.07
LA50 / LV32	0.99 ± 0.09	0.40 ± 0.02	0.42 ± 0.05	0.60 ± 0.04	0.42 ± 0.06
TA60	1.13 ± 0.13	1.05 ± 0.02	–0.24 ± 0.05	0.08 ± 0.10	–0.21 ± 0.05
TA63	0.30 ± 0.10	0.31 ± 0.02	–0.05 ± 0.04	–0.02 ± 0.34	–0.15 ± 0.14

Comparison of different mutations at the same positions in order to analyse the contributions of individual parts of each side-chain.

the transition state can be isolated (Fersht *et al*., 1992; Serrano *et al*., 1992a) (Table 5). The Φ_F for composite mutation Val → Ala32 is higher than that for the composite mutation Ile → Val32, arguing that the C^{δ} methyl group is less structured (i.e. interacting with other atoms to a lesser degree) in the transition state than the two C^{γ} methyl groups. We cannot correlate this with the structure of the native protein, since the crystal structure has a Leu and not an Ile at this position. For the mutation Val → Ala32, the interaction of $C^{\gamma 1}$ and $C^{\gamma 2}$ with other residues in the transition state, as measured by $\Phi_F H_2O$, is unchanged, within error, by the presence of other mutations such as at positions 38 and 50. In the case of the mutation Leu → Ala50, Φ-values vary between 0.23 and 0.42, depending on the presence of mutations at positions 32 and 38. Thus, these mutations subtly affect the structure formation at position 50 in the transition state.

The Φ-values for mutations in the minicore are not significantly larger or smaller than those at other sites in the protein. The minicore is not, therefore, a nucleation centre for folding *via* hydrophobic clustering, nor do tertiary interactions appear to form later than secondary interactions; tertiary interactions form in parallel with secondary structure interactions. The minicore interactions in the transition state are consistent with the overall structure of the transition state of folding of CI2 as an expanded loose version of the native state.

α-Helix

Folding of the α-helix of CI2 was probed with mutations at 12 of the 13 sites. Three of the α-helix residues, Ala16, Val19 and Ile20, are involved in the packing of the α-helix against the β-sheet to form the major hydrophobic core, and are described in detail under the hydrophobic core. The side-chains of the other sites mutated in the α-helix are all solvent-exposed, and make interactions almost exclusively with other helix residues only. They are all charged groups, with the exception of Leu21. Mutations were made to Ala and to Gly, at Ser12 (the N cap), Lys17, Lys18, Leu21, Gln22 and the C-cap, Lys24. All these mutations result in fractional Φ_F-values. At positions 14 and 15, the Glu side-chain has been mutated to Gln and to Asp, both of which are conservative mutations. There is, however, considerable inconsistency in the values of Φ_F for the different mutations at each of these two positions, which may indicate that these mutations cause some structural reorganisation in the folded protein, or that the folding / unfolding pathway is altered. However, it is more likely that the Φ_F-values obtained for the mutations to Gln and to Asp are not very reliable because of the relatively small values of $\Delta\Delta G_{U\text{-}F}$ ($\leqslant$0.5 kcal / mol). Mutation to Asn at these positions had a larger effect on the equilibrium free energy of unfolding, and the Φ_F-values of 0.72 (for Glu → Asn14) and 0.52 (for Glu → Asn15) are, therefore, a more reliable measure of the degree of structure formation at these sites.

A generally benign probe for solvent-exposed positions in α-helices is the comparison of Ala and Gly (''Ala → Gly scanning'', Matthews & Fersht, 1995). The difference in stability of Ala *versus* Gly appears to depend solely on the solvent-accessible surface area (Serrano *et al*., 1992b,c). Application to the helix here gives data in good agreement with the other probes (Table 6). The structure in the transition state becomes progressively weaker towards the C-cap, which itself is almost completely unstructured in the transition state. We see a gradation, though somewhat irregular, of Φ_F along the helix. It is clear that the N-cap box (residues 12, 14 and 15) is one of the most ordered regions of the protein in the transition state for folding, with Φ_F-values greater than 0.5.

β-Sheet

The structure of the β-sheet in the transition state of refolding for CI2 has been probed at six different sites, namely Thr3 (β-strand 1), Ile30 (β-strand 3), Val34 (β-strand 3), Ala58 (β-strand 5), Val60 (β-strand 6) and Val63 (β-strand 6). These residues were chosen because their side-chains interact primarily with other β-sheet residues in the native state. Since the mutation Ile → Val30 has only a very small effect on the equilibrium free energy of unfolding (it stabilises the protein by 0.08 kcal/mol), it was not subjected to Φ-value analysis.

Mutations in β-strands 1, 5 and 6 (residues 3, 58, 60 and 63) show very low Φ_F-values, indicating that these strands are not structured in the transition state. Indeed, these strands comprise the only part of CI2 where there is a very clear absence of native structure in the transition state. These residues constitute the N and C termini of the protein, and in the unfolded state they are likely to be very far apart. It is not surprising, therefore, that the coming together of the two separate ends of the protein constitutes a very late stage of the folding process. Further, the two C-terminal strands of the β-sheet of CI2 lack the completely regular hydrogen-bonding network of the canonical β-sheet. Water molecules form bridging hydrogen bonds between strands 4, 5 and 6, for example between the carbonyl oxygen atoms of Gln59 and Val60 and the amide nitrogen of Phe50 (McPhalen & James, 1987). In contrast, strand 4 is hydrogen-bonded in a regular fashion to strand 3, to form a relatively long stretch of parallel β-sheet.

The mutation Thr → Val3 displays a Φ_F-value of 0.52, which is considerably higher than those of the other two mutations made at position 3. However, when we combine the data we have at this site to construct the composite mutations Val → Ala3 and Ala → Gly3, we obtain low Φ_F-values (0.07 and –0.19) similar to the mutations Thr → Ala3 and Thr → Gly3. The anomalous behaviour of the Thr → Val mutation may be a result of a hydrophilic/hydrophobic switch involved which introduces a complication in the interpretation of Φ_F-values, since the possibility arises that the mutated site will be solvated in a manner different from that of the wild-type residue, leading to a different free energy of solvation. There is no general solution to this problem. It is apparent that more consistent results in terms of Φ_F-values are achieved when we focus on mutations where hydrophobic groups are removed, e.g. Val → Ala3 and Thr → Ala3. This pattern is repeated at residues 34 and 60, where the mutations Val → Thr (and, in the case of residue 34, Thr → Ala) have Φ_F-values which are significantly higher than those observed for mutations involving hydrophobic deletions. The $C^{\gamma 1}$ and $C^{\gamma 2}$ atoms of these residues are not completely buried and the solvation energies are again likely to contribute to the deviation in Φ_F-values. Val63, on the other hand, is completely buried and Φ_F for the mutation Val → Thr63 is in good agreement with those for the other mutations at this site. On the other hand, Ile30 has well-exposed $C^{\gamma 1}$ and $C^{\gamma 2}$ atoms, yet the Val → Thr30 mutation does not deviate from the other mutations at residue 30 in terms of Φ_F-values. Obviously it is not straightforward to generalize the impact of Val↔Thr mutations on Φ_F-value interpretation.

The central residues of β-strands 3 and 4 interact with the α-helix to form the major hydrophobic core of CI2. While the side-chain of Ile30 in β-strand 3 points away from the core and up towards the loop, it remains within 4.5 Å of the side-chain atoms of several core residues like Ile29 and Val47. Indeed, Φ_F-values for mutations at this position are higher than those for mutations in strands 1, 5 and 6. Judging from the fine-structure analysis of the secondary mutant Ala → Gly30 (Table 6), the C^{β} atom is less structured than the rest of the side-chain atoms, with a Φ_F-value of 0.15.

Val34 at the very edge of β-strand 4 appears to be structured to a similar degree to Ile30 in the transition state: Val → Ala34 displays a Φ_F-value of 0, but higher values, comparable to those at position 30, are obtained from unfolding studies and from a fine-structure analysis of the mutant Ala → Gly34 (Table 6).

Reactive site loop

The reactive site loop consists of ten residues between Gly35 and Ile44. This loop is very solvent-exposed, and mutations do not destabilize the protein to any great extent, except for that of Val38 whose side-chain takes part in tertiary interactions with Leu32 and Phe50 to form the hydrophobic minicore. Access of water to the site of mutation in conjunction with low values of $\Delta\Delta G_{U\text{-}F}^{H_2O}$ make Φ_F-value analysis difficult. Thr36 has been subjected to extensive mutagenesis, but only Thr → Val36 is sufficiently destabilized to yield a reliable value of Φ_F, of 0.19 (mutations Thr → Ser36 and Thr → Ala36 have insignificant effect on both $\Delta\Delta G_{U\text{-}F}^{H_2O}$ and $\Delta\Delta G_{\ddagger\text{-}U}^{H_2O}$). However, the mutation involves a hydrophilic/hydrophobic switch, and the interpretation of this figure is subject to the same limitations as discussed in the case of the β-sheet mutants. The mutation removes the hydrogen bond between $O^{\gamma 1}$ of Thr36 and the backbone O of Pro33, an interaction which does not appear to be very strong (our unpublished results). Thus, the hydrogen bond is hardly formed at all in the transition state of folding.

We have also performed a double-mutant cycle on the two residues Arg43 and Arg45. According to the crystal structure of pseudo-wild-type CI2, the two side-chains are aligned so that N^{H2} and N^{ϵ} of Arg43 can form hydrogen bonds with $O^{\delta 2}$ and $O^{\delta 1}$, respectively, of Asp45. In addition, there is extensive packing interaction between the two residues. Yet the interaction energy between the two residues is

very low (0.15 kcal / mol), and prevents us from calculating values of Φ_F^{int} with any precision. The values of Φ_F for the individual mutants are low for Arg → Ala43 and Arg → Ala43 / Asp → Ala45 (less than 0.1), but considerably higher for Asp → Ala43 alone (0.39). Presumably, this reflects interactions with other residues, namely Ile44, Glu26 or Arg46. The mutations Tyr → Gly42 and Tyr → Ala42 also have low value of Φ_F (0.07).

Turns

Lys2 is not part of a turn, but its side-chain N^{ζ} forms a hydrogen bond with $O^{\epsilon 2}$ of Glu7 (part of the type III turn between β-strand 1 and the α-helix). The interaction with Glu7 is very weak (0.07 kcal / mol) (Table 7), and so we cannot measure Φ_F^{int}. Thus, the free energy of refolding and unfolding for the double mutant Lys → Ala2 / Glu → Ala7 is simply the sum of two independent contributions from the mutations Lys → Ala2 and Glu → Ala7.

Lys2 is also close to the side-chain of Asp23, and this interaction is stronger, namely 0.58 kcal / mol, though no hydrogen bond is implicated in the pseudo-wild-type crystal structure. Asp23 appears to be involved in a non-native interaction during the transition state for the folding of CI2, since the mutant Asp → Ala23 folds considerably faster than wild-type, although the mutation destabilizes the protein overall by 0.96 kcal / mol. This non-native interaction probably involves Lys2, since the mutation Lys → Ala2 also refolds faster than wild-type, and has a value of Φ_F of –0.19. Probing the interaction energy between Lys2 and Asp23 also gives a negative value of Φ_F^{int} (–0.17). The negative Φ_F-values are reflected in both the folding and the unfolding data (except for Asp → Ala23). Interestingly, the mutant Lys → Met2, which retains the long non-polar side-chain but removes the polar hydrogen-bonding N^{ζ} atom, has a value of Φ_F of 0, indicating that any hydrogen bond involving Lys2 is completely absent in the transition state of refolding. Asp23 and Lys2 may be involved in a non-native interaction during refolding, which does not involve the hydrogen bond found in the native state.

Glu7 by itself is fairly structured in the transition state, with a Φ_F-value of 0.40, significantly larger than those for mutation of the surrounding residues Pro6 and Leu8 (Pro → Ala6 and Leu → Ala8 have Φ_F-values of 0.07 and 0.15, respectively). This may reflect early interactions with Trp5 and Glu4.

The residues in the β-hairpin loop between strands 4 and 5, as probed by mutations of Asp52 and Asn56, are almost completely unstructured in the transition state for refolding. Such low values of Φ_F were also observed in β-strands 5 and 6. Thus the C-terminal 13 residues from Asp52 to Gly64 only attain native structure after the transition state.

Interestingly, the mutant Asp → Asn52, whose equilibrium free energy of unfolding is identical to that of wild-type, refolds with a significantly lower rate constant than wild-type, giving $\Delta\Delta G_{\ddagger\text{-}U}^{H_2O}$ = 0.24(±0.02) kcal / mol. At 4 M GdmCl, the unfolding rate constant is identical to wild-type (Table 2); however, since the value of m_{ku} for Asn52 is slightly lower than the value for wild-type, Asn52 unfolds more slowly at higher values of GdmCl. Mutants with stability identical to wild-type may thus be affected differently during the folding pathway.

Similarly, Lys → Ala11 stabilizes CI2 by 0.42 kcal / mol, yet increases $\Delta G_{\ddagger\text{-}U}^{H_2O}$ by 0.21 kcal / mol. This is the opposite effect to that seen for Lys2 and Asp23, where mutations destabilize the protein but increase folding. All other mutations that have little effect on $\Delta\Delta G_{U\text{-}F}^{H_2O}$ (Gln → Ala22, Thr → Ala36, Thr → Ser36, Ile → Ala37, Thr → Asp39, Met → Leu40, Met → Leu40 / Lys → Met53) have positive fractional Φ-values (Tables 1 to 3).

Table 7. Ala → Gly scanning for the α-helix and β-sheet of CI2

Mutation	$\Delta\Delta G_{U\text{-}F}^{H_2O}$[a] (kcal mol^{-1})	$\Delta\Delta G_{\ddagger\text{-}F}^{4M}$[b] (kcal mol^{-1})	$\Delta\Delta G_{\ddagger\text{-}U}^{H_2O}$[c] (kcal mol^{-1})	$1 - \Phi_U^{4M}$	$\Phi_F^{H_2O}$
α-*Helix*					
AG17	1.84 ± 0.06	0.73 ± 0.03	0.74 ± 0.05	0.60 ± 0.02	0.40 ± 0.03
AG18	1.20 ± 0.15	0.30 ± 0.04	0.60 ± 0.04	0.75 ± 0.05	0.50 ± 0.07
AG22	0.58 ± 0.12	0.31 ± 0.03	0.19 ± 0.04	0.46 ± 0.12	0.33 ± 0.10
AG24	2.54 ± 0.12	2.20 ± 0.02	0.54 ± 0.06	0.14 ± 0.04	0.21 ± 0.02
β-*Sheet*					
AG30	1.40 ± 0.09	1.38 ± 0.03	0.27 ± 0.06	0.02 ± 0.07	0.19 ± 0.04
AG34	1.80 ± 0.12	0.99 ± 0.04	0.41 ± 0.05	0.45 ± 0.04	0.23 ± 0.03
AG60	1.72 ± 0.08	1.85 ± 0.02	0.17 ± 0.05	–0.07 ± 0.05	0.10 ± 0.03
AG63	1.99 ± 0.09	1.98 ± 0.02	0.06 ± 0.05	0.01 ± 0.05	0.03 ± 0.02

The data are for Gly *versus* Ala at each position, where the rate and equilibrium data for the Ala mutant are used in the previous equations as ''wild-type''.

[a] Change in the free energy of unfolding.

[b] Change in the free energy of the transition state of unfolding in 4 M GdmCl.

[c] Change in the free energy of the transition state of folding in water.

Discussion

Necessity and reliability of the protein engineering procedure

There are no stable intermediates that can be detected on the folding pathway of CI2 and so the only structure to be analysed is the transition state†. The only way of analysing transition states experimentally is by kinetics; and the only procedure for extracting structural information at the level of individual residues in the transition state is the use of kinetic measurements on mutants. There is thus currently only one procedure available for the experimental study of structural events on the folding pathway of CI2: the protein engineering method. The structure of the transition state for the folding and unfolding of CI2 has been so mapped by comparing the kinetics and equilibria of folding of a large number of mutants. There is satisfying agreement between the values of Φ_F measured directly from folding kinetics in water and Φ_F calculated from $1 - \Phi_U$ measured from unfolding kinetics in GdmCl solutions and those values extrapolated to water. This shows that the structure of the transition state is the same when measured in the directions of unfolding and refolding. Multiple mutations at the same site, mutations in contiguous regions, and the use of more advanced procedures such as double mutant cycles and the comparison of Ala and Gly (Ala → Gly scanning) at the same positions as the single mutations also give highly consistent data. We are confident, therefore, that the results of the Φ-value analysis are not artifactual but are satisfactorily probing the energetic changes and, by inference, the structural properties of the transition state. A further facet of Φ-value analysis is that it provides measurements of changes in energy that can be used to compare experiment with theory. The combined data are summarised in Table 8. It is worth emphasizing that the values of Φ_F are measured directly from reforming rate constants in water and do not require extrapolation from measurements in the presence of denaturants.

Gross nature of the transition state of folding and unfolding

Although values of Φ_F of 0 and 1 correspond to completely denatured and completely folded structures, respectively, there is not, in general, a linear relationship between a fractional value of Φ and the extent of formation of non-covalent bonds. There should, however, be an approximately linear relationship for the special case of the mutation of larger hydrophobic side-chains to smaller ones (Matouschek *et al*., 1989, 1990; Fersht *et al*., 1992). Many of the mutations in this study are of that type so Φ in those cases is a good indication of the extent of structure formation. Fractional Φ-values may result from genuinely weakened interactions or the reaction proceeding by parallel pathways, some of which have the element of structure tested by the Φ-value fully formed and others fully unfolded (or a mixture of the different processes), as discussed by Fersht (1995). A test has been devised to distinguish between the two mechanisms which, when applied to the folding of CI2, indicated genuine weakened interactions (Fersht *et al*., 1994).

The structure of the transition state is, to a first approximation, a relatively uniformly expanded form of the folded structure. This may be illustrated by a classical linear free energy plot: a plot of $\Delta\Delta G_{‡\text{-}F}$ *versus* $\Delta\Delta G_{U\text{-}F}$ (Figure 3) is approximately linear, with slope 0.7. This is equivalent to a mean value of Φ_U of 0.7 or $\Phi_F = 0.3$. This means that, on average, 70% of the free energy of the interactions is lost on reaching the transition state for unfolding from the folded state. This relatively uniform behaviour may be contrasted with that of the transition state for the unfolding of barnase: the plot of $\Delta\Delta G_{‡\text{-}F}$ *versus* $\Delta\Delta G_{U\text{-}F}$ (Figure 3) is scattered but has a set of points clustering around the line of slope 1, i.e. regions that are completely unfolded ($\Phi_U = 1$); another set around the line of slope 0, i.e. regions that are full-folded ($\Phi_U = 0$); and others values in between. The mean value of Φ_U is similar, however, at 0.69.

The 70% loss of energy on going from the folded state of CI2 to the transition state may be compared with a increase of 40% in the average degree of exposure (calculated from $m_{‡\text{-}F} / m_{U\text{-}F}$), and an increase of 25% in ΔC_P (Jackson & Fersht, 1991b), which is usually taken as a measure of the change in the surface exposure of hydrophobic residues.

Local variations in formation of structure

The α-helix shows a gradation of structure along its length with the N-cap >50% formed in terms of energy, and the C-cap almost completely unstructured. Thus, formation of the α-helix is on-going throughout the folding reaction and starts from the N-cap. The structure of the α-helix in the transition state may be a looser form of that in the folded protein, with the tertiary interactions that are likely to stabilize this secondary structure in the folded protein by packing against the β-sheet to form the core similarly weakened (see section on the hydrophobic core). A plot (not shown) of $\Delta\Delta G_{‡\text{-}F}$ *versus* $\Delta\Delta G_{U\text{-}F}$ for the helical residues is quite linear, however, with slope 0.65($\pm$0.08), correlation coefficient 0.87, indicating that the helix, when judged as a whole, is in the process of being formed during the transition state.

† The transition state for protein folding has characteristics similar to those in simple chemistry. Probing the energy surfaces of protein folding transition states by structure-reactivity relationships of physical-organic chemistry reveals behaviour consistent with their being at a saddle point (Matouschek & Fersht, 1993; Matthews & Fersht, 1995; Matouschek *et al*., 1995).

Table 8. Summary of the sites of mutation in CI2 and their Φ-values

Residue	Description	Solvent-accessible area (Å²)			Buried area (Å²)			Φ_F[a] (for mutation)
		Total	Side-chain	Non-polar side-chain	Total	Side-chain	Non-polar side-chain	
Whole protein		4183	3352	2398	6884	4730	3823	
A. *Core*[b]	The core consists of 12 residues with hydrophobic side-chains; Trp5, Leu8, Ala16, Val19, Ile20, Ala27, Ile29, Val47, Leu49, Val51, Ile57 and Pro61	133	81	81	1888	1420	1393	
Leu8[b]	On edge of hydrophobic core, in a type III reverse turn between Trp5 and Leu8, and also a type II reverse turn between Leu8 and Lys 11. C^{γ}, $C^{\delta1}$, and $C^{\delta2}$ atoms make contact with side-chains of Trp5, Ala16, Val19, Ile20, Pro61 and Lys11. Leu → Ala deletes interactions with α-helix and β-strand 6	6	6	6	174	131	131	0.2 (L → A)
Ala16	α-Helix residue, in centre of hydrophobic core. Completely buried. C^{β} makes contact with side-chains of Leu8, Leu49, and Ile57 and with backbone of Val13 and Glu15. Ala → Gly deletes interactions with N-terminal part of α-helix and β-strand 4	0	0	0	113	67	67	1.1 (A → G)
Val19[b]	α-Helix residue, on edge of hydrophobic core. Side-chain partially exposed. Interacts with Trp5, Leu8, Ile19 and Asp23. Val → Ala deletes interactions between β-strand 1 and the α-helix	62	59	59	98	58	58	–0.3 (V → A)
Ile20[b]	α-Helix residue, in centre of hydrophobic core. Completely buried. $C^{\delta1}$ methyl group packs against Trp5, Leu8, Ala16, Val47, Leu49 and Pro61. Ile → Val deletes interactions between α-helix and β-strands 1, 4 and 6	0	0	0	182	140	140	0.4 (I → V)
Ile29[b]	Ile29 is second residue of β-strand 3, at the edge of the core. Ile → Val deletes contacts of $C^{\delta1}$ methyl group with side-chains of Lys17, Ile20 and Leu21, removing interactions between β-strand 2 and α-helix. Ile → Ala deletes contacts between $C^{\gamma1}$ and $C^{\gamma2}$ methyl groups and side-chains of Ala27, Val31, Val47, and Leu49, removing interactions from β-strand 3 with α-helix and β-strand 4	27	5	5	155	135	135	0.2 (I → V) 0.3 (I → A)
Val47[b]	On edge of hydrophobic core but completely buried. First residue in β-strand 4. Val → Ala deletes contacts with Trp5, Ile20, Lys17, Ala27, Leu49, Pro61 and Val63, removing contacts with β-strand 1, middle of α-helix and β-strand 6	0	0	0	160	117	117	0.2 (V → A)
Leu49[b]	In centre of hydrophobic core, central residue of β-strand 4. Completely buried. Leu → Ala deletes interactions with residues Val13, Ala16, Ile20, Ile29, Val31, Val47, Ile57 and Pro61, removing interactions with α-helix, and β-strands 3, 4, 5 and 6	0	0	0	180	137	137	0.5 (L → A)
Val51[b]	Last residue in β-strand 4, at edge of core. Val → Ala deletes interactions with Val13, Val31, Leu47, Asp55 and Ile57, removing interactions with α-helix and β-strands 3, 4 and 5	13	11	11	147	106	106	0.3 (V → A)
Ile57[b]	Second residue in β-strand 5, at edge of core. $C^{\delta1}$ makes contacts with Val13, Ala16, Leu47 and Val51, the $C^{\gamma1}$ and $C^{\gamma2}$ methyl groups make contacts with Leu8, Val9, Ala16, Leu49 and Pro61. Ile → Val deletes interactions between β-strand 5 and α-helix and β-strand 4, Ile → Ala removes interactions with β-strand 1	3	0	0	179	140	140	0.1 (I → A)
Pro61	Second residue of β-strand 6 (O H-bonds to N^{H} of Trp5). Side-chain points into centre of core. Side-chain atoms contact Trp5, Leu8, Ile20, Val47, Arg48, Leu49, Ile57, Glu59 and Arg62, all in the core except Arg48, Glu59 and Arg62. Pro → Ala mutation removes contact with β-strands 4 and 5	2	0	0	141	105	105	0.0 (P → A)

continued overleaf

Table 8. *continued*

Residue	Description	Solvent-accessible area (Å^2) Total	Side-chain	Non-polar side-chain	Buried area (Å^2) Total	Side-chain	Non-polar side-chain	Φ_F[a] (for mutation)
B. *α-Helix*	The α-helix consists of 13 residues between Ser12 and Lys24.	809	700	503	1307	843	711	
Ser12	N-cap of the α-helix. Solvent-accessible surface mainly C^{β} (36 Å2). Interacts mainly with residues 32 to 34. Ser → Ala deletes interactions at the N-cap with Lys11, Val13, Glu14 and Glu15. Ser → Gly removes further interactions of the C^{β} methylene with these residues	48	44	36	74	3	8	0.4 (S → A) 0.3 (S → G)
Glu14	(N-cap + 2) position of the α-helix. Side-chain points away from helix into solvent. Solvent-accessible surface area mainly C^{γ} (31 Å^2), C^{β} (27 Å^2) and $O^{\epsilon 1}$ (23 Å^2). Glu → Ala removes interactions with Ser12 and Val13 (for triple mutant SG12/EA14/EA15)	98	94	66	85	45	−5	1.2 (E → Q) 0.2 (E → D)
Glu15	(N-cap + 3) position of the α-helix. Side-chain directs towards N-cap of α-helix, Solvent-accessible surface mainly $O^{\epsilon 1}$ (31 Å^2), C^{γ} (20 A^2) and $O^{\epsilon 2}$ (17 Å^2). Glu → Ala removes interactions with Lys11, Ser12 and Glu14 (for triple mutant SG12/EA14/EA15)	89	83	35	94	55	26	0.5 (E → Q) 0.5 (E → D)
Lys17	(N-cap + 5) position of the α-helix. Side-chain points into solvent. Solvent-accessible surface entirely side-chain atoms, mainly N^{ζ} (42 Å^2) and C^{δ} (20 Å^2). Lys → Gly removes interactions with Val13, Glu14, Glu15, Ala16, Lys18, Ile20, all in the α-helix, and with Ile29 in β-strand 3	72	72	31	139	95	88	0.3 (K → A) 0.4 (K → G) 0.4 (A → G)
Lys18	Central position of the α-helix. Side-chain points into solvent. Highly solvent-accessible side-chain, mainly C^{ϵ} (51 Å^2), C^{δ} (34 Å^2), N^{ζ} (37 Å^2) and (14 Å^2). Lys → Gly removes interactions within the α-helix at Glu14, Glu15, Lys17 and Val19	148	140	104	63	27	15	−0.4 (K → A) 0.7 (K → G) 0.5 (A → G)
Leu21	(C-cap − 3) position of the α-helix. Solvent-accessible surface mainly side-chain, viz $C^{\delta 1}$ (49 Å^2), C^{β} (11 Å^2) and $C^{\delta 2}$ (10 Å^2). Leu → Ala removes interactions within the helix at Lys17, Ile20, Lys24 and Pro25, with Ala27, and with Ile29 in β-strand 3. Leu → Gly removes interactions of the C^{β} methylene within the α-helix at Lys17, Lys18, Ile20 and Gln22	68	63	63	112	74	74	0.3 (L → A) 0.4 (L → G)
Gln22	(C-cap − 2) position of the α-helix. Side-chain points away from helix into solvent. Solvent-accessible surface mainly $N^{\epsilon 2}$ (57 Å^2), C^{γ} (25 Å^2), O (20 Å^2) and $O^{\epsilon 1}$ (19 Å^2). Gln → Gly removes interactions with Lys18, Val19, Leu21 and Asp23, all in the α-helix	149	124	38	39	20	15	0.1 (Q → G)
Asp23	(C-cap − 1) position of the α-helix. Side-chain points towards Trp5. Side-chain within 4.5 Å of Trp5, Lys2, Val19 and Gln22. $O^{\delta 2}$ within hydrogen-bonding distance of $N^{\epsilon 1}$ of Trp5 (2.80 Å), but no direct hydrogen bonds to Lys2 (closest distance: 4.41 Å between Lys2 N^{ζ} and Asp23 $O^{\delta 2}$). 79 Å^2 solvent-accessible surface area, mainly $O^{\delta 2}$ (35 Å2), $O^{\delta 1}$ (5 Å^2) and O (28 Å^2). Asp → Ala removes interactions with Lys2 (N-terminal part of CI2), Trp5 and Val19 (both in core) and α-helix	79	47	40	79	66	4	−0.3 (D → A) 0.3 (A → G)
Lys24	The C-cap of the α-helix. Side-chain points towards end of reactive site loop (residues 62 to 64). Does not form salt-bridge with neighbouring Glu26. Solvent-accessible surface virtually entirely side-chain, mainly N^{ζ} (18 Å^2) and C^{γ} (13 Å^2). N^{ζ} H-bonds to $O^{\epsilon 1}$ of Glu26. This H-bond is removed on mutation to Gly. Lys → Gly also removes interactions with Trp5 in β-strand 1, with Ile20, Asp23 in the helix, with Pro25, Glu26, Ala27 in a reverse turn following the C-cap of the helix, Ile44, Asp45, Arg46, Val47 in β-strand 4, and Val63	41	39	21	170	128	98	−0.4 (K → A) 0.1 (K → G) 0.2 (A → G)

C. β-*Sheet*	1280	1105	658	3740	2621	2255		
Thr3	First residue of β-strand 1 (CO H-bonds to NH of Val63 on β-strand 4). Side-chain oriented towards Val63. $C^{\gamma 2}$ and C^{β} contact Val63 (β-strand 6), while $O^{\gamma 1}$ contacts Lys2 and Glu4 (β-strand 1). Thr → Val replaces $O^{\gamma 1}$ with $C^{\gamma 2}$, and Thr → Ala removes part of contact to β-strand 6	98	76	52	48	26	22	0.1 (T → A) 0.5 (T → V) 0.1 (T → G)
Ile30	Third residue of β-strand 3 (NH H-bonds to Val47, CO H-bonds to Leu49, both on β-strand 4). Side-chain points away from core. Solvent -accessible surface mainly $C^{\delta 1}$ (45 Å^2) and $C^{\gamma 2}$ (17 Å^2). C^{β} contacts Ile48, Val47 and Arg48, $C^{\gamma 1}$ contacts Gly28, Ile29 and Arg46, $C^{\gamma 2}$ contacts Val31, Leu32 and Arg48 while $C^{\delta 1}$ contacts Ile29. All these residues are in β-strands 3 and 4. Ile → Val, Ile → Ala and Ile → Gly remove contacts to β-strand 3 and β-strand 4, while Ile → Thr changes interaction from non-polar to polar in this region	54	50	50	128	90	90	0.3 (I → A) 0.3 (I → G) 0.4 (I → T) 0.2 (A → G)
Val34	Last residue of β-strand 2 (NH H-bonds to Val51 in β-strand 4). Just before reactive site loop (residues 35 to 44). Points towards side-chain of Ala58. Solvent-accessible surface area mainly O (17 Å^2), $C^{\gamma 1}$ (29 Å^2) and $C^{\gamma 2}$ (13 Å^2). C^{β} contacts Pro33, Gly35, Val51 and Glu59, $C^{\gamma 1}$ contacts Gly35 and Glu59 while $C^{\gamma 2}$ contacts Pro33, Phe50, Val51, Asp52, Ala58 and Glu59. All these residues except Gly35 (active loop) and Asp52 (β-turn) are in β-strands 3 to 5. Val → Ala and Val → Gly remove these contacts, while Val → Thr changes a non-polar interaction to a polar interaction in these regions	59	42	42	101	75	75	0 (V → A) 0.2 (V → G) 0.2 (V → T) 0.2 (A → G)
Ala58	Last residue of β-strand 5 (NH H-bonds to Phe50). C^{β} points towards side-chain of Val34, and contacts Val34, Phe50, Val51, Asp52, Asn56, Ile57 and Glu59 in β-strands 3 to 5. Ala → Gly removes these contacts	37	12	12	76	55	55	0.1 (A → G)
Val60	First residue of β-strand 6. Side-chain points towards Pro6. Solvent -accessible surface mainly $C^{\gamma 2}$ (36 Å^2) C^{β} contacts Gly64, Arg62, Glu59, Pro6, and Val9, $C^{\gamma 1}$ contacts Glu4, Trp5, Pro6 and Pro61 while $C^{\gamma 2}$ contacts Pro6, Val9 and Glu59. All these residues except Pro6 are in β-strands 1,2 and 6. Val → Ala and Val → Gly remove these contacts and Val → Thr changes a non-polar interaction to a polar one in this region	60	51	51	100	66	66	0 (V → A) 0 (V → G) 0.1 (A → G)
Val63	Penultimate residue of β-strand 6 (NH H-bonds to Thr3). Side-chain entirely buried. Side-chain points towards Thr22. C^{β} contacts Thr3, Ile45, Arg46, Arg62 and Gly64, while $C^{\gamma 1}$ contacts Lys24, Tyr42, Arg43 and Gly64 and $C^{\gamma 2}$ contacts Lys2, Trp5, Lys24, Ile44, Arg46, Val47 and Arg62. These contacts are in β-sheets 1,3 and 6, the reactive loop and two turn regions. Val → Ala and Val → Gly remove contacts with all these regions. Val → Thr changes a non-polar interaction to a polar one	7	0	0	154	117	117	0.1 (V → T) 0 (V → A) 0 (V → G) 0 (A → G)
D. *Minicore*		133	106	106	425	323	323	
Leu32	Antepenultimate residue of β-strand 3 (NH H-bonds to CO of Leu49, CO H-bonds to Val51). Side-chain points away from β-sheet towards reactive site loop. Only side-chain is solvent-accessible, mainly $C^{\delta 2}$ 35 Å^2). Side-chain atoms mainly pack to the two other minicore side-chains (Val38 and Phe50), but also contact Ile30, Val31, Pro33 (β-strand 3), Thr36 (loop) and Leu49 (β-strand 4 and core). Leu → Ile, Leu → Val and Leu → Ala modify and reduce contact to these regions	43	42	42	137	95	95	0.2 (L → A) −0.3 (L → I) −0.1 (L → V)
Val38	Two residues before scissile bond in reactive loop. Side-chain points from loop down to β-sheet. Side-chain solvent-accessible surface area mainly $C^{\gamma 2}$ (28 Å^2). Side-chain contacts to minicore as well as Thr36, Ile37, Thr39 (loop) and Arg48 (β-strand 4) are removed by the mutation Val → Ala	58	33	33	102	84	84	0.1 (V → A)

continued overleaf

Table 8. *continued*

Residue	Description	Solvent-accessible area (Å^2)			Buried area (Å^2)			Φ_F[a] (for mutation)
		Total	Side-chain	Non-polar side-chain	Total	Side-chain	Non-polar side-chain	
Phe50	Penultimate residue in β-strand 4 (CO H-bonds to NH of Ala58). Side-chain points up from β-sheet towards the reactive site loop. Only its side-chain is solvent-accessible, mainly $C^{\epsilon 1}$ (17 Å^2), $C^{\delta 1}$ (6 Å^2) and C^{ζ} (5 Å^2). Side-chain contacts to minicore and also to Pro33, Val34 (β-strand 3), Thr36 (loop), Arg48 (β-strand 4), Leu49, Val51 and Glu59 (core). Phe → Leu, Phe → Val and Phe → Ala (which successively remove the ring atoms) change these packing interactions	32	31	31	186	144	144	0.3 (F → L, F → V, F → A)
E. *Turns & loop*								
Lys2	In coil at N terminus just before β-strand 1. All side-chain atoms interact with Trp5. N^{ζ} just above the pyrrole moiety of the indole atom. All except C^{β} interact with Glu4. In addition, C^{δ}, C^{ϵ} and N^{ζ} interact with Asp23 (a hydrogen bond to $O^{\delta 2}$ of Asp42 may be mediated by a bridging water molecule). C^{ϵ} and N^{ζ} interact with Glu7. N^{ζ} is within hydrogen-bonding distance of $O^{\epsilon 2}$ of Glu7 (2.62 Å). C^{β} and C^{γ} interact with Thr3. Lys → Met maintains interactions with all the residues (β-strand 1, α-helix and turn before helix), but removes the H-bonds to Glu7 and Asp23 and the polar interaction with Trp5, while Lys → Ala removes all the interactions	168	80	44	43	87	75	–0.2 (K → A) 0 (K → M)
Pro6	In type III turn between β-strand 1 and α-helix. Side-chain points into solvent and contacts with Glu4 and Trp5 (β-strand 1), Glu7 (turn before helix) and Val60 (β-strand 6). Pro → Ala removes contact with β-strands 1 and 6 and the turn	75	65	65	68	41	41	0.1 (P → A)
Glu7	In type III turn betwen β-strand 1 and α-helix. Side-chain points back along the backbone towards the N terminus and contacts Lys2, Glu4, Trp5, Pro6 and Leu8. $O^{\epsilon 2}$ hydrogen-bonds to N^{ζ} of Lys2. Glu → Ala removes interactions with β-strand 1, turn before helix and the hydrogen bond to Lys2. Glu → Gln prevents deprotonation at low pH, but should otherwise not perturb interactions	94	66	56	89	72	5	0.4 (E → A)
Glu26	In type I reverse turn between α-helix and β-strand 3. Side-chain sticks straight into solvent and has contacts to Lys24 (helix), Pro25, Ala27 (same turn) and Asp45 (β-strand 4). $O^{\epsilon 1}$ H-bonds to N^{ζ} of Lys24. Glu → Ala removes these interactions, including the H-bond to Lys24	157	132	56	26	6	5	0.4 (E → A)
Thr36	At N-terminal part of reactive loop. Side-chain points from loop down towards the β-sheet. C^{β} interacts with Leu32, Pro33, G135, Ile37 and Phe50, $O^{\gamma 1}$ H-bonds to O of Pro33. This hydrogen bond appears to stabilize the segment of the reactive site loop composed of residues 34 and 35, which partly projects out. $O^{\gamma 1}$ also interacts with Leu32 and Gly35, while $C^{\gamma 2}$ interacts with Leu51, Ile56 and Va157. Thr → Val, Thr → Ser and Thr → Ala remove the H-bond to Pro52, and change interaction with loop and minicore	54	36	31	92	66	43	0.2 (T → V)
Ile37	In beginning of reactive loop. Highly exposed, mainly side-chain. Side-chain atoms contact backbone of Gly33, Thr36 and Val38 (all in loop). Ile → Ala removes contact with reactive loop only	163	141	141	19	–1	–1	

Thr39	Residue before scissile residue. Side-chain points down towards the β-sheet. Side-chain solvent-accessible surface mainly $C^{\gamma 2}$ (67 Å^2). Side-chain atoms contact side-chain of Glu41 and backbone of Met40 and Val38. Hydrogen-bond between $O^{\gamma 1}$ of Thr39 and $O^{\gamma 1}$ of Glu51. Thr → Ala and Thr → Asp remove this H-bond and a few contacts with reactive loop	94	82	78	52	20	–4	
Met40	Scissile residue. Hyperexposed. Side-chain atoms contact backbone of Thr39 and Glu41. Met → Ala removes contact with reactive loop only	199	162	162	5	–2	–2	
Glu41	Residue after scissile residue. Side-chain points down towards β-sheet. $O^{\epsilon 2}$ H-bonds to N^{H1} of Arg46. Other side-chain atoms contact Thr39 ($O^{\gamma 1}$ H-bonds to $O^{\gamma 1}$ of Thr39), Met40, Tyr42 and Arg43. Glu → Ala removes H-bond to Arg46 and contacts with reactive loop, including the H-bond	80	62	50	103	76	11	
Arg43	At C-terminal part of reactive loop, where loop makes turn into β-strand 3. End of side-chain solvent exposed, mainly N^{H1} (51 Å^2), Cδ (26 Å^2) and N^{H2} (23 Å^2). C^{γ} interacts with Glu41, Tyr42 and Arg46 while C^{δ}, N^{ϵ}, C^{ζ}, N^{H1} and N^{H2} all interact with Asp45 only. Only interactions with reactive loop (mainly Asp45, but also Glu41, Tyr42 and Arg46). N^{ϵ} H-bonds to $O^{\delta 1}$ of Asp45, possibly stabilizing turn between loop and β-strand 3. Arg → Ala removes these interactions	89	87	40	152	109	49	0.1 (R → A)
Asp45	Last residue in the loop. Solvent-accessible surface mainly C^{β} (26 Å^2). $O^{\delta 1}$ H-bonds to N^{ϵ} of Arg43. Other side-chain atoms interact with side-chain of Arg43 and backbone of Ile44, Glu26 and Arg46. Asp → Ala removes these interactions	49	44	33	102	63	15	
Asp52	In turn between β-strand 4 and β-strand 5. Solvent-accessible surface area mainly $O^{\delta 2}$ (25 Å^2) and C^{β} (19 Å^2). Side-chain only contacts Lys53, Leu54 (same turn), Asp55 and Asn56 (β-strand 5). $O^{\delta 2}$ interacts closely with the NH groups of Lys53, Leu54, Asp55 and Asn26, stabilizing the β-hairpin turn. In addition, C^{γ} and $O^{\delta 1}$ interact with Asp55 and Asn56. Asp → Ala removes these interactions, including the H-bonding network, while Asn → Ala only affects protonation at lower pH	54	45	21	97	61	27	0.1 (D → A)
Asn56	First residue of β-strand 5. Side-chain points away into solvent. Solvent -accessible surface area mainly $N^{\delta 2}$ (27 Å^2) and $O^{\delta 1}$ (11 Å^2). C^{γ} interacts with Ile57, Asp55, Lys11, Asp52, Leu54 and Gly10, $O^{\delta 1}$ interacts with NH on Asn56, Asp55, Lys11, Ser12, Leu54 and Gly10 while $N^{\delta 2}$ interacts with Asp52, Leu54 and Gly10. Side-chain contacts Gly10, Lys11 (turn), Ser12 (helix N-cap), Leu54, Asp55 (turn) and Ile57 (β-strand 5). Asn → Ala removes these interactions, while Asn → Asp introduces a charge	41	41	4	117	72	40	0.2 (N → D) 0.1 (N → A)

Area of extended surface (i.e. the completely unfolded protein) calculated according to values from Miller *et al*. (1987). In this study, accessible surface area of each amino acid residue was calculated for residue X in in a Gly-X-Gly tripeptide, with the main-chain in an extended conformation (Miller *et al*., 1987).

[a] Values of Φ_F are measured from refolding rate constants that are directly measured in water.

[b] From Jackson *et al*. (1993b).

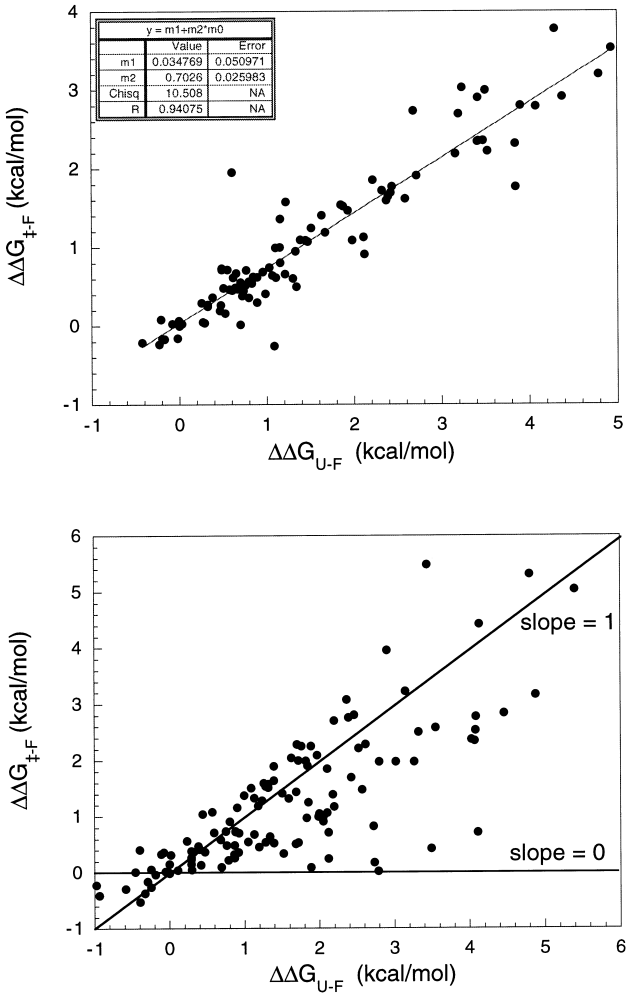

Figure 3. Plots of $\Delta\Delta G_{‡\text{-}F}$ *versus* $\Delta\Delta G_{U\text{-}F}$ for the unfolding in water of (top) CI2 and (bottom) barnase. The linear regression line for the data for CI2 is given. Lines of slope 0 and 1 are indicated for barnase.

The degree of formation of structure is very low in the β-sheet. Φ-Value analysis allows us to monitor formation of interactions of side-chains, and is, therefore, not a direct measure of the extent of secondary structure formation, which rather depends on backbone-backbone hydrogen-bonding. However, secondary structure formation is a prerequisite for the side-chains to be able to establish contacts within the β-sheet. The side-chains of the residues probed in the β-sheet, namely Thr3, Ile30, Val34, Ala58, Val60 and Val63, all interact primarily with other β-sheet side-chains. Their structure in the transition state thus reports on the overall β-sheet structure.

In β-strand 3, there is a similar extent of formation of structure in the transition state for positions 34 and 30. We observe similar values at position 29, at the edge of the core. In strand 4, Φ_F-values at position 43 (end of reactive site loop), 47 (edge of core), 50 (minicore) and 51 (edge of core) are close in magnitude to each other and to those in strand 3. It is evident that the folding process shows no clear differentiation between core residues and non-core residues in the β-strands 3 and 4. Rather, the data give excellent support to the model that the main body of CI2 folds as a single co-operative unit, with secondary and tertiary interactions forming concomitantly with few exceptions, namely at residues 49 and 16. Strands 1, 5 and 6 appear to be completely unfolded in the transition state. Despite the formation of the β-sheet requiring that residues far apart in sequence (at opposite ends of the molecule) be brought in close proximity, their alignment is not part of the rate-limiting step of the folding process, and once again demonstrates that formation of intact secondary structure is not a prerequisite for further rearrangements in folding.

The three-residue hydrophobic patch in the reactive site loop, which we call the minicore, has been thoroughly examined by double mutant cycles. All the kinetic data converge to suggest that the minicore is ca 20 to 40% structured in the transition state relative to the folded state. Further, it is formed along a single folding pathway, rather than parallel pathways (Fersht *et al*., 1994). This conclusion is reached both by mutation of individual residues, double mutant cycle analysis and fine-structure analysis.

Relationship to structures of peptide fragments of CI2

In studies elsewhere, the structures of peptides corresponding to the progressive elongation of CI2 from its N terminus have been analysed (Prat Gay *et al*., 1995; unpublished work, this laboratory). The native secondary structural elements develop fully only in parallel with the formation of tertiary interactions (Prat Gay *et al*., 1995). Fragment (1 to 50) is not compact and requires elongation to (1 to 53) before compact structure can be detected. The helix does not form significantly until the protein is nearly intact, and even the fragment 1 to 60, which lacks only the last four residues, has fluctuating structure and some non-native interactions (Prat Gay *et al*., 1995). Small fragments which include the residues corresponding to the α-helix of the intact protein have been shown to have negligible structure in water, and in the helix-promoting solvent trifluoroethanol, only nascent helix is formed (Itzhaki *et al*., 1995). Titration of the formation of nascent helix with trifluoroethanol and extrapolation to 0 M trifluoroethanol does suggest, however, that the nascent helix is present to the order of ~3% in water. Further, the transition state for the association of two complementary fragments of CI2 (fragment 1 to 40 and fragment 41 to 64), analysed using the protein engineering method (Prat Gay *et al*., 1994; Ruiz-Sanz *et al*., 1995), is remarkably similar in structure to the transition state for folding of the intact protein. Therefore, the co-operative folded structure is assembled only as

the two fragments associate. The overwhelming conclusion that can be drawn from the study of fragments of CI2 is that isolated sequences of this protein cannot form significant fractions of native-like structure independently and, therefore, that tertiary interactions in this small protein are both crucial for the native fold and to the transition state for its formation. (See Creighton (1995) for an analysis of these phenomena in terms of co-operativity of interactions.)

Relationship to theoretical models: nucleation-condensation

The structure of the transition state for the folding of intact protein and the structures of fragments support a folding pathway for CI2 in which there is no early, independent, formation of native-like secondary structure. Rather, folding of CI2 is a very concerted reaction; secondary and tertiary structure are formed concomitantly and are consolidated throughout the folding reaction. These observations do not fit to a simple framework model in which extensively formed elements of structure are formed early and then coalesce, or even to any model that requires a pre-formed initiation site.

Based on the observation that the co-operativity of multiple interactions is required to stabilize both the native structure and the transition state for its formation, we and Creighton (1995) envisage the folding of CI2 to proceed qualitatively, as follows, in general terms. There is rapid random searching of conformations in the unfolded state of the protein under conditions that favour folding until sufficient tertiary interactions are made that stabilize certain elements of secondary structure. When sufficient interactions are made, the transition state is reached and there follows the rapid formation of the final structure.

Nucleation-condensation

Our data fit an even more specific model, which we term nucleation-condensation. This involves a nucleus that consists primarily of adjacent residues. The nucleus does not form stable structure without assistance from interactions made with residues that are distant in sequence. Formation of the small nucleus cannot be solely rate determining since a significant fraction of the overall structure must be in the approximately correct conformation to provide the long-range contacts to stabilize the nucleus. In CI2, for example, the α-helix has a very weak tendency to be formed in a nascent manner, driven by local interactions. The helix remains embryonic until sufficient long range interactions are built up that it becomes stable. The rest of the protein then condenses around the helix and the other native interactions that are developed during the stabilization of the helix. The onset of stability as the multiple interactions are made is so rapid that the helix is still in the process of being consolidated as the transition state is reached. Thus, formation of the nucleus is coupled with more general formation of structure and so nucleation is just part of the rate determining step. The coupling of nucleation and condensation leads to a relatively compact transition state. The whole of the molecule is thus involved in forming the transition state, the nucleus is simply the best formed part of the structure in the transition state.

This nucleation-condensation mechanism has certain characteristics. First, is that the nucleation site does not need to be extensively preformed in the denatured state. A theoretical analysis of the optimization of the rates of protein folding suggests that the less it is formed in the denatured state, the better (Fersht, 1995b). It may or may not be detectable in isolated peptides, but it is only essential that it is extensively formed in the transition state. Further, it need not be formed completely in the transition state but, because of the onrush of co-operative interactions that stabilize it, it may be in the process of being formed in the transition state. Second, although much of the structure is from local, contiguous residues, there are important stabilizing contributions from long range interactions, i.e. from contacts with residues that are far removed in sequence.

Abkevich *et al*. (1994) have performed Monte Carlo simulations of the folding of a 36-mer ''protein'' on a lattice. They observed in their simulations the formation of a specific nucleus of interactions whose formation constituted the ''transition state'' of the reaction and precipitated subsequent rapid folding, and the ''protein'' was found to fold *via* a nucleation growth mechanism†. Further, without prior knowledge of our experimental data, E. Shakhnovich and co-workers (E. Shakhnovich, personal communication) have

† A referee has pointed out that the nucleation-growth mechanism of Abkevich *et al*. (1994) is for the formation of a molten globule and not native protein. However, those authors now believe their proposal is more general and have given us permission to quote: ''In the original paper (Abkevich *et al*., 1994) it was suggested that the observed nucleation mechanism applies to formation of the ''native-like'' Molten Globule with the native backbone conformation already formed but with rotating side-chains. This assertion was based on the hypothesis that the rate-limiting step in folding is associated with tight-packing of side-chains while backbone conformation is already close to native. The existing experimental data do not support this hypothesis. Rather they suggest that the rate-limiting step is the formation of the backbone fold, probably with simultaneous packing of side-chains. In this case, the nucleation mechanism of the formation of backbone fold found in lattice models applies to the whole process of folding. This is especially clear in the case when folding is kinetically two-state, as in the case of CI2'' (E. Shakhnovich, personal communication).

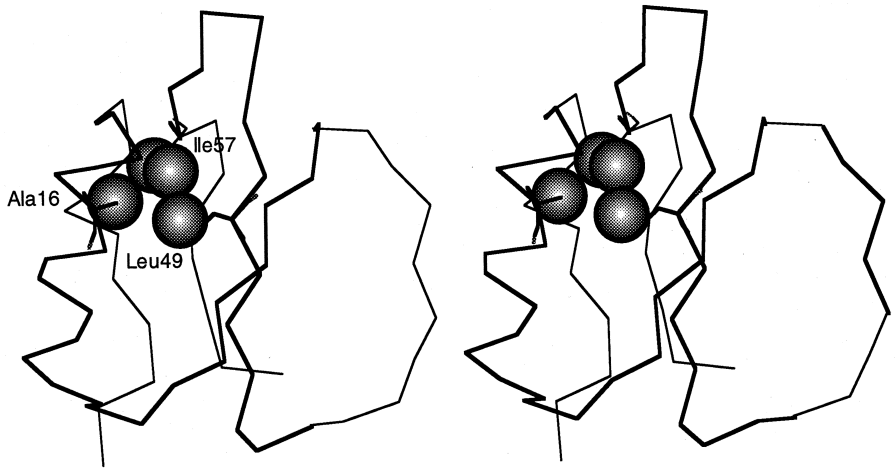

Figure 4. Stabilization of the formation of the N-terminal region of the α-helix in the transition state by the interaction of the side-chain of Ala16 with residues that will form the hydrophobic core. The filled spheres (C^{β} of Ala16; $C^{\delta 1}$ and $C^{\gamma 1}$ of Ile57; and $C^{\delta 1}$ of Ile57) are drawn with the full van der Waals' radii. The N-terminal region of the helix and the ancillary interactions constitute the proposed nucleation site.

predicted that Ala16 is a key residue in nucleation (Figure 4). The overall agreement between simulation and experiment is so encouraging that experiment and theory could well combine to give a satisfactory computer model for folding.

It is now common practice to search for initiation sites in the folding of proteins by searching for structure in isolated peptides by experiment or computation (Hirst & Brooks, 1995). The results of the experiments on CI2 and the lattice simulations suggest that this may be an inconclusive process since the nucleation sites do not necessarily emerge until the transition state is reached. Further, even if significant structure is found, it does not mean that it is a nucleation site or that it is formed at the earliest stages of folding.

Concerted *versus* stepwise folding mechanisms in general

Barnase is larger than CI2, comprising separate modules (Yanagawa *et al*., 1993), and folds *via* a distinct folding intermediate in which the regions containing the major α-helix and β-sheet are reasonably well formed and connected by a loose and weak hydrophobic core (Fersht, 1993). In the absence of each other, i.e. in peptide fragments, the structures are formed only weakly under folding conditions (Kippen *et al*., 1994a) and have only very weak structure under denaturing conditions (Kippen *et al*., 1994b). Further, destabilizing mutations in the helix cause it to be less folded in the intermediate and the subsequent transition state for folding, and the extent of formation of the overall

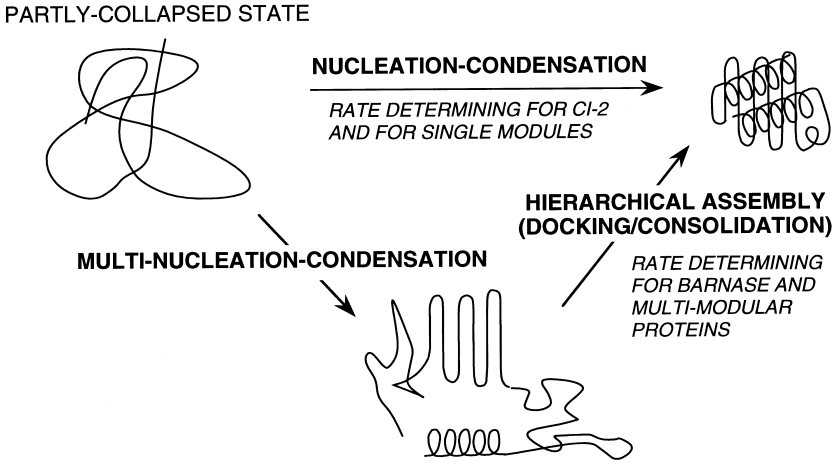

Figure 5. A unified scheme for the initiation of folding in CI2 and barnase, extending the mechanism presented by Otzen *et al*. (1994) using the additional data and analysis in this study. The ''partly-collapsed'' state is the ''unfolded'' state under conditions that favour folding. CI2, which folds as a single co-operative unit, folds in one step, which is the nucleation-condensation process. The folding of barnase is consistent with the formation of a folding intermediate by nucleation steps and then the structure consolidating in the rate determining step. Although the folding of barnase is drawn as stepwise, the nucleation-condensation and docking-consolidation processes do tend to merge under certain conditions (M. Oliveberg, P. Dalby and A.R.F., unpublished).

structure is coupled to the stability of the helix (Matthews & Fersht, 1995). This shows that the folding intermediate of barnase is also formed in a concerted manner. Thus barnase fits also a nucleation-condensation folding mechanism (Figure 5). We suggest that the folding of CI2 is a model for the folding of modules in larger proteins.

Whether or not folding is a concerted or stepwise process depends on the stability of individual substructures of the protein when they are considered in isolation or loosely complexed with each other. For example, since the α-helix and β-sheet of CI2 are unstable in the absence of the rest of the protein, they are formed in a concerted manner. If they were stable, then the protein could form by a framework or collision / diffusion model. The same is true for individual modules within a protein. If a module is stable in the absence of the other modules, then it could form independently. The more the individual structures are stable, the more stepwise and hierarchical the folding process. We thus speculate that small proteins tend to fold by nucleation-condensation mechanisms but larger proteins will have a tendency to form in their later stages by a hierarchical assembly of smaller units, but the folding of the small units will tend to be nucleation-condensation processes (Figure 5).

Materials and Methods

Materials

All chemicals are as described in previous papers (Jackson & Fersht, 1991a; Jackson *et al*., 1993a). Mutagenesis, expression and purification of the wild-type and mutant proteins are as described previously (Jackson *et al*., 1993a). The buffer used in the equilibrium and kinetic folding experiments was 2-(*N*-morpholino)ethanesulfonic acid (Mes) purchased from Sigma. A 1 M stock solution of Mes (pH 6.25) contained 415 mM of the free acid and 585 mM of the sodium salt. 16 novel mutant proteins are presented in this paper: Lys → Met2, Lys → Ala2, Glu → Ala7, LysA → Ala2 / Glu → Ala7, Lys → Ala2 / Asp → Ala23, Ala → Gly16, Asp → Ala23, Thr → Ala36, Tyr → Ala42, Tyr → Gly42, Arg → Ala43, Asp → Ala45, Arg → Ala43 / Asp → Ala45, Asp → Ala52, Asp → Asn52, Pro → Ala61.

GdmCl equilibrium denaturation experiments

The intrinsic fluorescence of CI2 increases on denaturation, allowing unfolding and refolding to be monitored by fluorescence spectroscopy. The maximal change in fluorescence upon denaturation is obtained with an excitation wavelength of 280 nm and an emission wavelength of 356 nm. All experiments were performed at 25°C.

A Perkin Elmer LS5B luminescence spectrometer was used for the GdmCl equilibrium denaturation experiments, with a bandpass of 10 nm. The GdmCl solutions were prepared using a Hamilton Microlab M robot by aliquoting appropriate volumes of a solution of 7.5 M GdmCl containing 50 mM Mes (pH 6.25) and a solution of 50 mM Mes (pH 6.25). For each data point in the denaturation experiment, 100 μl of CI2 stock solution in 50 mM Mes (pH 6.25) were diluted into 800 μl of the appropriate denaturant concentration, using a SMI positive displacement pipetter. Final concentrations were 2.5 μM CI2 and 50 mM Mes (pH 6.25); final concentrations of GdmCl were 0 M to 5.5 M, in 0.1 M or 0.2 M increments. The protein-denaturant solutions were equilibrated at 25°C for approximately one hour before measuring their fluorescence. Many experiments were repeated with GdmCl solutions that had been prepared in the conventional manner by weighing the appropriate amount for each individual concentration into a volumetric flask. Identical results were obtained.

Kinetic unfolding 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 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 thermostatted separately using two water baths and the temperature of each was monitored using an Edale Instrument Thermometer and maintained to ±0.1 deg.C. Data were acquired with a Tandon Target Microcomputer, a DT2801 Data Translation board and the Bio-kine data acquisition software package (Bio-logic), and analysed using the programme Kaleidagraph (Abelbeck Software). Unfolding was initiated by rapidly diluting one volume of the protein solution (approximately 20 μM) in 50 mM Mes (pH 6.25) with ten volumes of concentrated GdmCl solution (containing 50 mM Mes, pH 6.25). This resulted in final GdmCl concentrations of between 3 and 7 M. The lowest denaturant concentration was chosen to result in at least 98% unfolded protein. An excitation wavelength of 280 nm and an emission wavelength of 356 nm were used. Excitation and emission slit bandwidths of 10 nm were used.

Kinetic refolding experiments

Refolding reactions above 0.5 M GdmCl were monitored using the same apparatus set-up as for unfolding experiments. Refolding was initiated by rapid 1:11 dilution of unfolded protein (approximately 20 μM) in 6.5 M GdmCl and 50 mM Mes (pH 6.25) into different concentrations of GdmCl solution in 50 mM Mes (pH 6.2). For lower final concentrations of GdmCl, an Applied Photophysics Stopped-Flow Spectrophotometer Model SF 17MV was used. The temperature of the cell and reservoir syringes was maintained by thermostatting with a Grant LTD6 water bath. Temperatures were maintained to ±0.1 deg.C based on an internal temperature probe in the stopped flow apparatus which had previously been calibrated against an Edale Instrument Thermometer. Protein was denatured for these studies by addition of HCl to a concentration of 20 mM (pH 1.7). Refolding was initiated by rapid mixing of one volume of the protein with one volume of 100 mM Mes (pH 6.65) (consisting of 21.5 mM of the free acid and 78.5 mM of the sodium salt). The final pH of the solution was pH 6.25, 50 mM Mes, and the final protein concentration was approximately 10 μM. Refolding was performed in the absence and in the presence of low concentrations (up to 0.6 M, with 0.1 M increments) of GdmCl. An excitation wavelength of 280 nm was used, and slit bandwith of 2 nm. A glass

cut-off filter was used to allow emission above 315 nm to be monitored. Refolding rates obtained from the two different sets of refolding experiments were identical within error when unfolded protein was refolded into GdmCl concentrations that could be monitored by both the Perkin-Elmer MPF-44B fluorescence spectrophotometer and the Applied Photophysics Stopped-Flow Spectrophotometer.

Acknowledgements

We thank Eugene Shakhnovich for prior communication of results about Ala16 and for clarifying comments about nucleation-growth mechanisms.

References

Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1994). Specific nucleus as the transition state for protein folding: Evidence from the lattice model. *Biochemistry*, **32**, 10026–10036.

Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell*, **38**, 835–840.

Clarke, J. & Fersht, A. R. (1993). Engineered disulfide bonds as probes of the folding pathway of barnase – increasing the stability of proteins against the rate of denaturation. *Biochemistry*, **32**, 4322–4329.

Creighton, T. E. (1995). Protein-folding – an unfolding story. *Curr Biol.* **5**, 353–356.

Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993). Cooperativity in protein-folding kinetics. *Proc. Natl Acad. Sci. USA*, **90**, 1942–1946.

elMasry, N. F. & Fersht, A. R. (1994). Mutational analysis of the N-capping box of chymotrypsin inhibitor-2. *Protein Eng.* **4**, 777–782.

Fersht, A. R. (1993). Protein folding and stability – the pathway of folding of barnase. *FEBS Letters*, **325**, 5–16.

Fersht, A. R. (1995a). Characterizing transition-states in protein-folding – an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84.

Fersht, A. R. (1995b). Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA*, in the press.

Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.

Fersht, A. R., Itzhaki, L. S., Matthews, J. M. & Otzen, D. E. (1994). Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA*, **91**, 10426–10429.

Hammond, G. S. (1955). A correlation of reaction rates. *J. Am. Chem. Soc.* **77**, 334–338.

Harpaz, Y., elMasry, N. F., Fersht, A. R. & Henrick, K. (1994). Direct observation of better hydration at the N-terminus of an α-helix with glycine rather than alanine as the N-cap residue. *Proc. Natl Acad. Sci. USA*, **91**, 311–315.

Harrison, S. C. & Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl Acad. Sci. USA*, **82**, 4028–4030.

Hirst, J. D. & Brooks, C. L. I. (1995). Molecular-dynamics simulations of isolated helices of myoglobin. *Biochemistry*, **34**, 7614–7621.

Horovitz, A. & Fersht, A. R. (1990). Strategy for analysing the cooperativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol.* **214**, 613–617.

Horovitz, A. & Fersht, A. R. (1992). Co-operative interactions during protein folding. *J. Mol. Biol.* **224**, 733–740.

Horovitz, A., Serrano, L. & Fersht, A. R. (1991). COSMIC analysis of the major α-helix of barnase during folding. *J. Mol. Biol.* **219**, 5–9.

Itzhaki, L. S., Neira, J.-L., Prat Gay, G. de, Ruiz-Sanz, J. & Fersht, A. R. (1995). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. *J. Mol. Biol.* **254**, 289–304.

Jackson, S. E. & Fersht, A. R. (1991a). Folding of chymotrypsin inhibitor-2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.

Jackson, S. E. & Fersht, A. R. (1991b). Folding of chymotrypsin inhibitor-2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry*, **30**, 10436–10443.

Jackson, S. E. & Fersht, A. R. (1994). Contribution of residues in the reactive-site loop of chymotrypsin inhibitor-2 to protein stability and activity. *Biochemistry*, **33**, 13880–13887.

Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. & Fersht, A. R. (1993a). The effect of cavity creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry*, **32**, 11262–11269.

Jackson, S. E., elMasry, N. & Fersht, A. R. (1993b). 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. *Biochemistry*, **32**, 11270–11278.

Karplus, M. & Weaver, D. C. (1976). Protein folding dynamics. *Nature*, **260**, 404–406.

Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics – the diffusion-collision model and experimental data. *Protein Sci.* **3**, 650–668.

Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.

Kippen, A., Sancho, J. & Fersht, A. R. (1994a). Structural studies on peptides corresponding to mutants of the major alpha-helix of barnase. *Biochemistry*, **33**, 3778–3786.

Kippen, A., Arcus, V. L. & Fersht, A. R. (1994b). Folding of barnase in parts. *Biochemistry*, **33**, 10013–10021.

Kraulis, P. J. (1991). Molscript, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crysallog.* **24**, 946–950.

Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **85**, 44–45.

Li, A. & Daggett, V. (1994). Characterization of the transition-state of protein unfolding by use of molecular-dynamics – chymotrypsin inhibitor-2. *Proc. Natl Acad. Sci. USA*, **91**, 10430–10434.

Ludvigsen, S., Shen, H., Kjaer, M., Madsen, J. C. & Poulsen, F. M. (1991). Refinement of the three-dimensional solution structure of barley serine proteinase inhibitor 2 and comparison with the structures in crystals. *J. Mol. Biol.* **222**, 621–635.

Matouschek, A. & Fersht, A. R. (1993). Application of

physical organic chemistry to engineered mutants of proteins – hammond postulate behavior in the transition state of protein folding. *Proc. Natl Acad. Sci. USA*, **90**, 7814–7818.

Matouschek, A., Kellis, J. T., Jr, Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, **342**, 122–126.

Matouschek, A., Kellis, J. T., Jr, Serrano, L., Bycroft, M. & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature*, **346**, 440–445.

Matouschek, A., Serrano, L. & Fersht, A. R. (1992). The folding of an enzyme. 4. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 819–835.

Matouschek, A., Matthews, J. M., Johnson, C. M., & Fersht, A. R. (1994). Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions. *Protein Eng.* **7**, 1089–1095.

Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., & Fersht, A. R. (1995). Movement of the position of the transition state in protein folding. *Biochemistry*, **34**, in the press.

Matthews, J. M. & Fersht, A. R. (1995). Exploring the energy surface of protein-folding by structure-reactivity relationships and engineered proteins – observation of Hammond behaviour for the gross structure of the transition-state and anti-Hammond behaviour for structural elements for unfolding of barnase. *Biochemistry*, **34**, 6805–6814.

McPhalen, C. A. & James, M. N. G. (1987). Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry*, **26**, 261–269.

Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656.

Nanzer, A. P., Poulsen, F. M., van Gunsteren, W. F. & Torda, A. E. (1994). A reassessment of the structure of chymotrypsin inhibitor-2 (CI-2) using time-averaged NMR restraints. *Biochemistry*, **33**, 14503–14511.

Otzen, D. E., Itzhaki, L. S., elMasry, N. F., Jackson, S. E. & Fersht, A. R. (1994). Structure of the transition state for the folding / unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl Acad. Sci. USA*, **91**, 10422–10425.

Otzen, D. E. & Fersht, A. R. (1995). Side-chain determinants of β-sheet stability. *Biochemistry*, **34**, 5718–5724.

Pace, C. N. (1986). Determination and analysis of urea and guanidinium denaturation curves. *Methods Enzymol.* **131**, 266–279.

Prat Gay, G. de, Ruiz-Sanz, J., Davis, B. J. & Fersht, A. R. (1994). The structure of the transition state for the association of 2 fragments of the barley chymotrypsin inhibitor-2 to generate native-like protein and implications for mechanisms of protein-folding. *Proc. Natl Acad. Sci. USA*, **91**, 10943–10946.

Prat Gay, G. de, Ruiz-Sanz, J., Neira, J. L., Itzhaki, L. S. & Fersht, A. R. (1995). Folding of a nascent polypeptide chain *in-vitro* – cooperative formation of structure in a protein module. *Proc. Natl Acad. Sci. USA*, **92**, 2683–3686.

Ptitsyn, O. B. (1973). Stage mechanism of the self-organization of protein molecules. *Dokl. Acad. Nauk.* **210**, 1213–1215.

Ptitsyn, O. B. (1991). How does protein synthesis give rise to the 3D-Structure. *FEBS Letters*, **285**, 176–181.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α-helices. *Science*, **240**, 1642–1648.

Ruiz-Sanz, J., Prat Gay, G. de, Otzen, D. E. & Fersht, A. R. (1995). Protein-fragments as models for events in protein-folding pathways – protein engineering analysis of the association of two complementary fragments of the barley chymotrypsin inhibitor-2 (CI-2). *Biochemistry*, **34**, 1695–1700.

Sali, A., Shakhnovich, E. & Karplus, M. (1994). Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* **235**, 1614–1636.

Santoro, M. M. & Bolen, D. W. (1992). A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry*, **31**, 4901–4907.

Serrano, L., Matouschek, A. & Fersht, A. R. (1992a). The folding of an enzyme. 3. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805–818.

Serrano, L., Neira, J. L., Sancho, J. & Fersht, A. R. (1992b). Effect of alanine versus glycine in α-helices on protein stability. *Nature*, **356**, 453–455.

Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R. (1992c). α-helix stability in proteins. 1. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by slycine or serine at solvent-exposed surfaces. *J. Mol. Biol.* **227**, 544–559.

Tanford, C. (1968). Protein denaturation. *Advan. Protein Chem.* **23**, 121–282.

Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl Acad. Sci. USA*, **70**, 697–701.

Yanagawa, H., Yoshida, K., Torigoe, C., Park, J. S., Sato, K., Shirai, T. & Go, M. (1993). Protein anatomy – functional roles of barnase module. *J. Biol. Chem.* **268**, 5861–5865.

Edited by J. Karn

(Received 16 August 1995; accepted 15 September 1995)