Strain in the folding nucleus of chymotrypsin inhibitor 2

Andreas G Ladurner, Laura S Itzhaki and Alan R Fersht

Background: Chymotrypsin inhibitor 2 (CI2) is a member of the class of fast-folding small proteins, which is very suitable for testing theories of folding. CI2 folds around a diffuse extended nucleus consisting of the single α helix and a set of hydrophobic residues. In particular, Ala16 has been predicted and independently found to interact with Leu49 and Ile57, hydrophobic residues that are highly conserved among homologues. We have characterised in detail the interactions between these residues in the folding nucleus of the protein by using double-mutant cycles.

Results: Surprisingly, we find that there is some destabilising strain in the transition state for folding of the wild-type protein between Ala16 and Ile57. Further, we find that the strain is larger in the native state of the protein. This is shown directly in the unfolding kinetics, which clearly show a release of strain. The net result of this is that the presence of both residues speeds up folding. Ala16 and Leu49 interact favourably in the transition state, but have no net interaction energy in the native state.

Conclusions: Part of the folding nucleus of the protein fits together more snugly in the transition state than it does in the native state. Interactions between some of the closely packed residues in the folding nucleus of CI2 may perhaps be optimised for the rate of folding and not for stability.

Address: MRC Cambridge Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge, CB2 2QH, UK.

Correspondence: Alan R Fersht
E-mail: arf10@cus.cam.ac.uk

Key words: catalysis, double-mutant cycles, nucleation site, protein engineering

Received: **08 July 1997**
Revisions requested: **06 August 1997**
Revisions received: **28 August 1997**
Accepted: **05 September 1997**

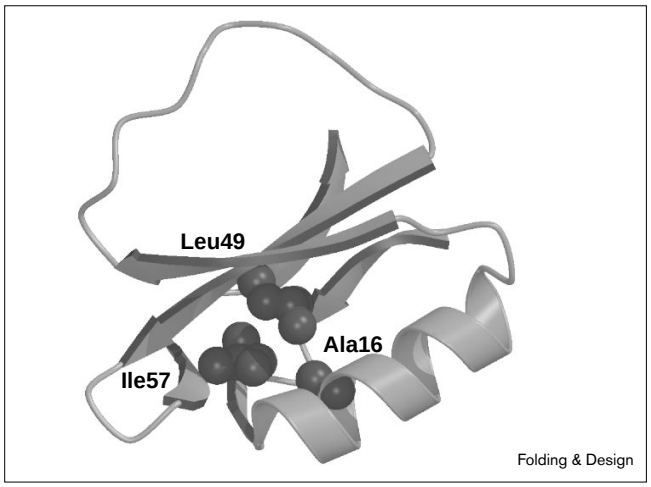
Published: **23 October 1997**
http://biomednet.com/elecref/1359027800200363

Folding & Design 23 October 1997, **2**:363–368

Introduction

The folding of the 64-residue protein chymotrypsin inhibitor 2 (CI2; Figure 1) conforms to a two-state model both at equilibrium and in its kinetics [1]. No intermediates accumulate, and there is only a single rate-determining transition state. CI2 is the first of a now large class of small fast-folding proteins which has attracted much interest from both experimentalists and theoreticians [2–15]. It has a single helix extending from residue 12 to 25. A nucleation-condensation model has been invoked for the folding pathway of CI2, in which the N-terminal region of the helix forms the core of a folding nucleus, which is stabilised in the transition state for folding by hydrophobic interactions with residues distant in primary sequence [16,17]. The helix is unstable in the absence of these hydrophobic interactions in small peptide fragments of CI2 [18] and is formed only in peptide fragments longer than 60 residues [19,20], thus emphasising the importance of long-range interactions in stabilising the protein.

Figure 1



Structure of CI2 showing the residues forming the nucleus in the transition state for folding of the protein. The coordinates were taken from the X-ray crystal structure [40]. The picture was generated using the programs Molscript [41] and Raster3D [42].

The degree of formation of native interactions in the transition state for folding has been measured by *Φ*-value analysis of mutated proteins ([21]; a *Φ*-value of 0 means that an interaction is as unformed in the transition state as in the denatured state, a value of 1 means that the interaction is as well formed as in the native state). All *Φ*-values for CI2 are fractional, apart from that of Ala16 (relative to the mutant Gly16) in the N-terminal part of the α helix, which is 1 [17]. The sidechain of Ala16 is buried in the hydrophobic core and interacts with the sidechains of three conserved core residues: Leu8, Leu49 and Ile57.

Shakhnovich and coworkers [12] independently predicted the importance of Ala16 in the folding nucleus using a computer algorithm based on sequence conservation and

lattice models of protein folding; Li and Daggett [22,23] have simulated the unfolding transition state of CI2 using molecular dynamics and find good agreement with the experimental results; further, Wolynes and coworkers [24] more recently found good agreement between the predicted transition state ensemble for CI2 and the one derived from experiment.

Amino acid residues in homologous protein sequences are often conserved for protein function, such as residues important for enzymatic catalysis. It has been found that more stable proteins can often be engineered by mutation of these residues but with a loss of activity [25–30]. There is thus an evolutionary compromise between optimal activity and stability. Further, catalysis of chemical reactions implies a selective stabilisation of the transition state relative to the ground states of a reaction.

We have measured the interaction energies between residues Ala16, Leu49 and Ile57 using double-mutant cycles [31]. Our kinetic measurements indicate that there is destabilising strain in the native state of the protein between Ala16 and Ile57. This strain is reduced in the transition state for folding. Thus, key interactions between residues involved in the formation of the folding nucleus are not optimised for the stability of the ground state. Rather, the site is designed for optimal interactions in the transition state. Upon consolidation of the structure to the native state, the interactions become strained.

Results

Double-mutant cycle analysis

The formation of sidechain interactions during protein folding pathways can be mapped by using the protein engineering approach [6,32,33]. A more sophisticated approach is the use of double-mutant cycles, which allow one to probe specific pairwise interactions [31,33,34]. A cycle involves mutating singly and then pairwise each residue in a pair X and Y that interact, and noting whether the effects of mutation are additive or not. A coupling energy, $\Delta\Delta G_{int}$, between sidechains of these residues is calculated from equation 1:

$$\Delta\Delta G_{int}^{D-N} = \Delta\Delta G_{E-XY\rightarrow E-Y}^{D-N} + \Delta\Delta G_{E-XY\rightarrow E-X}^{D-N} - \Delta\Delta G_{E-XY\rightarrow E}^{D-N} \quad (1)$$

where $\Delta\Delta G_{int}^{D-N}$ is the coupling energy in the native structure N relative to its absence in the denatured state D, $\Delta\Delta G_{E-XY\rightarrow E-Y}^{D-N}$ is the difference in the free energy of denaturation between the double mutant XY and single mutant Y, $\Delta\Delta G_{E-XY\rightarrow E-X}^{D-N}$ is the difference between the double mutant XY and single mutant X, and $\Delta\Delta G_{E-XY\rightarrow E}^{D-N}$ is the difference between the free energy of the double mutant XY and the wild-type protein. If the effect of a double mutant on the stability of the protein is equal to the sum of the effects of the two single mutants, then the effects of the two mutations are independent and the sidechains do not interact, as is often found [35]. If, however, $\Delta\Delta G_{E-XY\rightarrow E}^{D-N} \neq \Delta\Delta G_{E-XY\rightarrow E-Y}^{D-N} + \Delta\Delta G_{E-XY\rightarrow E-X}^{D-N}$, then the two residues do interact and the discrepancy is the coupling energy between them. This can occur when the residues interact with each other or the rate-limiting step of the reaction is changed. The coupling energy may be the result of direct or indirect interactions. $\Delta\Delta G_{int}^{\ddagger-D}$, the interaction energy in the transition state relative to the unfolded state, is calculated from the values of $\Delta\Delta G^{\ddagger-D}$ in a similar manner [33]. A Φ-value for the pairwise interaction can be calculated from the ratio of $\Delta\Delta G_{int}^{\ddagger-D}$ to $\Delta\Delta G_{int}^{D-N}$.

Kinetic unfolding measurements show better interactions between the mutated residues on going from the native state to the transition state

The double mutants unfold considerably faster than the wild-type protein (Table 1). Plots of ln k_u versus GdmCl

Table 1

Unfolding rates for wild-type CI2 and the hydrophobic core mutants.

Protein	ln k_u 2 M GdmCl (kcal mol^{-1})	$\Delta\Delta G^{\ddagger-N}$ 2 M GdmCl (kcal mol^{-1})	ln k_u 3 M GdmCl (kcal mol^{-1})	$\Delta\Delta G^{\ddagger-N}$ 3 M GdmCl (kcal mol^{-1})	ln k_u 4 M GdmCl (kcal mol^{-1})	$\Delta\Delta G^{\ddagger-N}$ 4 M GdmCl (kcal mol^{-1})
Wild type	−6.37 ± 0.28	–	−5.22 ± 0.03	–	−3.90 ± 0.02	–
A16G	−7.17 ± 0.43	+0.48 ± 0.30	−5.71 ± 0.23	+0.29 ± 0.14	−4.30 ± 0.09	+0.23 ± 0.05
L49A	−4.39 ± 0.09	−1.17 ± 0.17	−2.87 ± 0.04	−1.39 ± 0.03	−1.48 ± 0.01	−1.43 ± 0.01
I57A	−0.65 ± 0.19	−3.38 ± 0.20	+0.79 ± 0.03	−3.56 ± 0.03	+2.02 ± 0.02	−3.50 ± 0.02
A16G/L49A	−3.82 ± 0.02	−1.51 ± 0.16	−2.44 ± 0.01	−1.65 ± 0.02	−1.19 ± 0.01	−1.61 ± 0.01
A16G/I57A	+0.14 ± 0.04	−3.85 ± 0.17	+1.51 ± 0.03	−3.98 ± 0.03	+2.73 ± 0.04	−3.93 ± 0.03
I57V	−6.53 ± 0.13	+0.09 ± 0.18	−5.30 ± 0.10	+0.05 ± 0.06	−3.56 ± 0.05	−0.20 ± 0.03

Unfolding rate constants (ln k_u) were determined in the following ranges of denaturant: wild type, 5–7 M guanidinium chloride (GdmCl); A16G, 4.5–7 M GdmCl; L49A, 3.5–7 M GdmCl; I57A, 3–5 M GdmCl; A16G/L49A, 2.5–6.5 M GdmCl; A16G/I57A, 1.75–6 M GdmCl and I57V, 5–7 M GdmCl. $\Delta\Delta G$s are determined by $\Delta\Delta G = RT$ (ln k_u' − ln k_u), where k_u' and k_u are the unfolding rate of the wild-type protein and mutant protein, respectively, at 2 M, 3 M and 4 M GdmCl. Wild type, A16G and I57V do not show any curvature on unfolding within the range of GdmCl concentrations tested, so extrapolation to 2 M, 3 M and 4 M GdmCl is reasonably accurate. The other mutants show some curvature, so measurements at 4 M GdmCl are the most accurate.

Table 2

Coupling energies between sidechains on unfolding.

Interaction	$\Delta\Delta G^{\ddagger-N_{int}}$ 2 M (kcal mol^{-1})	$\Delta\Delta G^{\ddagger-N_{int}}$ 3 M (kcal mol^{-1})	$\Delta\Delta G^{\ddagger-N_{int}}$ 4 M (kcal mol^{-1})
Ala16–Leu49	0.81 ± 0.38	0.54 ± 0.14	0.41 ± 0.06
Ala16–Ile57	0.94 ± 0.40	0.71 ± 0.14	0.66 ± 0.06

The coupling energies are for the interactions in the transition state relative to the native state. $\Delta\Delta G^{\ddagger-N_{int}}$ 2 M, $\Delta\Delta G^{\ddagger-N_{int}}$ 3 M and $\Delta\Delta G^{\ddagger-N_{int}}$ 4 M were calculated from the unfolding rates in 2 M, 3 M and 4 M GdmCl, respectively (Table 1). Positive coupling energies are favourable.

concentration are linear for the wild-type protein and most mutants [17], but there is curvature in such plots for highly destabilised mutants (M. Oliveberg, Y.J. Tan, M. Silow and A.R.F., unpublished observations). Extrapolation of unfolding data to 0 M [GdmCl] may lead to some error, so we have calculated double-mutant cycles for 2 M, 3 M and 4 M [GdmCl], over which range we can measure the unfolding rate constants of the mutants directly (or with a very small extrapolation) and extrapolate the data on wild-type CI2 from measurements in the range 5–7 M [GdmCl].

The calculated coupling energies between residues Ala16 and Leu49 or Ile57, respectively, are positive (Table 2), indicating that the residues interact favourably in the transition state relative to the ground state. On going from the native state to the unfolding transition state, there are coupling energies between Ala16 and Leu49 of 0.41 ± 0.06 kcal mol^{-1} and between Ala16 and Ile57 of 0.66 ± 0.06 kcal mol^{-1} (at 4 M GdmCl). This shows directly an increase in interaction energy between these residues as the protein unfolds from its native state.

Kinetic folding measurements show there is strain between Ala16 and Ile57 in the transition state

The double mutants A16G/L49A and A16G/I57A fold very slowly (Table 3). The transition state for folding of the protein is destabilised by 2.91 ± 0.03 and 1.91 ± 0.01 kcal mol^{-1} in the two double mutants, respectively. A16G/L49A is the most slowly folding mutant analysed so far in the laboratory. Real-time NMR experiments on the refolding of this protein show that the appearance of native proton chemical shifts following a pH jump (see Materials and methods section) correlates very well with the fluorescence change of the Trp5 residue upon refolding; moreover all peaks appear with identical rates (A.G.L., T. Killick and A.R.F., unpublished observations).

$\Delta\Delta G_{int}^{\ddagger-D}$, the interaction energy between the mutated residues in the transition state of folding relative to the unfolded state, is calculated from the values of $\Delta\Delta G^{\ddagger-D}$, as described above [33]. Because these measurements are measured directly in water and do not require any extrapolation from denaturant, the protein folding rate constants we measure are very precise. Remarkably, there is an unfavourable interaction between Ala16 and Ile57 in the transition state for folding (Table 4), with a coupling energy of –0.38 ± 0.02 kcal mol^{-1}. Ala16 and Leu49, on the other hand, interact favourably in the transition state for folding, with a coupling energy of 0.31 ± 0.05 kcal mol^{-1}.

Equilibrium unfolding measurements show that there is strain in the native state

The thermodynamic parameters for the folding of the single mutants, A16G, L49A and I57A, have been reported previously [16,17,36,37]. The double mutants A16G/L49A and A16G/I57A are the most destabilised of all the mutants studied so far, resulting in losses of stability of 4.30 ± 0.13 and 5.66 ± 0.17 kcal mol^{-1}, respectively (Table 5).

The long extrapolation of equilibrium unfolding data to 0 M [GdmCl] leads to large errors. We therefore calculated the coupling energies at 3 M [GdmCl], where there is only a slight extrapolation (Table 6). The residues show unfavourable interactions in the native state. The coupling

Table 3

Folding rate constants for wild-type CI2 and mutants.

Mutant	k_f * (s^{-1})	$m^{\ddagger-D}$ (M^{-1})	$\Delta\Delta G^{\ddagger-D\ H_2O}$ (kcal mol^{-1})	$\Phi_F^{H_2O}$ †
Wild type	59 ± 1	–1.82 ± 0.12	0	0
A16G	8.3 ± 0.2	–2.23 ± 0.04	1.16 ± 0.01	1.07 ± 0.04
L49A	1.8 ± 0.1	–2.35 ± 0.12	2.06 ± 0.03	0.54 ± 0.02
I57A	32 ± 1	–2.10 ± 0.07	0.36 ± 0.02	0.08 ± 0.03
A16G/L49A	0.44 ± 0.02	–2.80 ± 0.11	2.91 ± 0.03	0.57 ± 0.02
A16G/I57A	2.37 ± 0.01	–3.22 ± 0.12	1.91 ± 0.01	0.31 ± 0.02
I57V	48 ± 1	–1.88 ± 0.02	0.12 ± 0.01	–‡

Kinetic measurements were performed at 25°C and pH 6.3, as described in detail in [17]. *The rate of folding, k_f, was measured by a pH jump from acid (20 mM HCl) to a final pH of 6.3. †The Φ-value, Φ_F, is derived from the following equation: $\Phi_F^{H_2O} = \Delta\Delta G^{\ddagger-D\ H_2O} / \Delta\Delta G^{N-D\ [D]_{50\%}}$, where $\Delta\Delta G^{\ddagger-D} = -RT \ln(k_f/k_f')$. ‡The Φ-values for I57V cannot be calculated accurately, since the changes in free energy are too small.

Table 4

Coupling energies between sidechains on folding.

Interaction	$\Delta\Delta G^{\ddagger-D}_{int}$ (kcal mol^{-1})
Ala16–Leu49	+0.31 ± 0.05
Ala16–Ile57	−0.38 ± 0.02

The coupling energies are for the interactions in the transition state relative to the denatured state. Calculated from the free energy change on folding ($\Delta\Delta G^{\ddagger-D\ H_2O}$) between wild-type, single-mutant and double-mutant proteins (data from Table 3). Positive coupling energies are favourable; negative coupling energies are unfavourable.

energies between Ala16 and Leu49 are –0.8 ± 0.4 kcal mol^{-1} and between Ala16 and Ile57 are –2.0 ± 0.2 kcal mol^{-1}.

We have also calculated the equilibrium values from the rate constants, as there is two-state folding. The direct equilibrium measurements for the coupling energy the residues agree favourably with the ones derived from kinetics (Table 6). The folding nucleus of CI2 would thus appear to fit together more snugly in the transition state than it does in the native state. The net result is that the native protein folds faster than do any of the mutants in which the sidechains have been deleted.

Φ-Values for Ala16→Gly in unstrained mutants

By calculating the energetics of making a second mutation in the single-mutant proteins, we can determine the energy cost of the mutations in the native state and transition state in the absence of strain. We use the mutant I57A as the parent protein (see Table 2), because the strain has been released in this mutant. The stabilisation energy of Ala relative to Gly at position 16 in mutant I57A is 2.0 kcal mol^{-1}, which is much greater than the value of 1.1 kcal mol^{-1} as determined from the single mutant A16G. The Φ-value for the mutation A16G is calculated to be 0.79 in I57A. This value is somewhat lower than the Φ-value for the mutation in the wild type (1.06). This overestimate of the Φ-value for A16G in the single mutants was also observed in the association of two complementary fragments of CI2 [20]. The Φ-values for the association of fragments (1–40) and (41–64) was 1.14 compared to 1.06 for the uncleaved protein.

Discussion

Extensive double-mutant cycle analysis of the major α helix [34], surface salt bridges [38,39], and the hydrophobic cores of barnase revealed no unfavourable interactions in the native state or in the major transition state for folding. The remarkable result in this paper is that the opposite is observed in the hydrophobic core of CI2 for those residues that form the folding nucleus. We find that on folding there are destabilising interactions between Ala16 and Ile57, while Ala16 and Leu49 interact favourably. During protein unfolding, we can detect a release of strain between Ala16 and Leu49 and Ala16 and Ile57. Our kinetic measurements thus clearly indicate that Ala16 and Ile57 interact unfavourably in the native state of the protein.

There is, thus, non-optimal packing of sidechains in the hydrophobic core of CI2 in the native state, consistent with an earlier finding that the mutation I57V actually stabilises the protein by 0.2 kcal mol^{-1} [36], when most other deletion mutations destabilise the protein. Its folding rate was marginally decreased relative to wild type, however (Table 3). Hence, there appears to be tension between the γ-methyl of the isoleucine sidechain and its neighbours, in particular Ala16. The crystal structure of I57V was determined [36] in order to investigate this unusual result, and a comparison of the structure with that of wild-type CI2 indicated that there is some repacking around the cavity created by the mutation [36]. Also, double-mutant cycle analysis of I29A/I57V shows that Ile57 and Ile29 interact unfavourably in the native state. Valine is more commonly found than isoleucine at the corresponding position in

Table 5

Equilibrium unfolding parameters.

Mutant	m_{D-N} (kcal mol^{-1} M^{-1})	$[D]_{50\%}$ (M)	$\Delta\Delta G^{D-N\ H_2O}$ * (kcal mol^{-1})	$\Delta\Delta G^{D-N\ 3M}$ † (kcal mol^{-1})
Wild type	1.90 ± 0.03	4.00 ± 0.01	0	0
A16G	1.80 ± 0.07	3.44 ± 0.02	1.41 ± 0.28	1.11 ± 0.04
L49A	1.98 ± 0.12	2.02 ± 0.03	3.60 ± 0.28	3.84 ± 0.12
I57A	1.93 ± 0.21	1.79 ± 0.05	4.15 ± 0.41	4.24 ± 0.26
A16G/L49A	2.38 ± 0.03	1.39 ± 0.01	4.30 ± 0.13	5.73 ± 0.06
A16G/I57A	2.45 ± 0.15	0.79 ± 0.02	5.66 ± 0.17	7.33 ± 0.35
I57V	1.82 ± 0.12	4.10 ± 0.05	0.14 ± 0.51	−0.10 ± 0.14

Changes in the free energies of unfolding upon mutation determined by reversible GdmCl denaturation experiments measured at 25°C in 50 mM MES, pH 6.3. *Variations in individual values of m_{D-N} are large; $\Delta\Delta G^{D-N\ H_2O} = m'\ [D]'_{50\%} - m\ [D]_{50\%}$, where m' and $[D]'_{50\%}$ are the parameters for the mutants and m and $[D]_{50\%}$ for the wild-type protein. †$\Delta\Delta G^{D-N\ 3M}$ is calculated at 3 M GdmCl concentration as above, but $[D]_{50\%}$ is replaced by ($[D]_{50\%}$ − 3).

Table 6

Coupling energies between the residues at equilibrium.

Interaction	At equilibrium* $\Delta\Delta G^{D-N}_{int\ 3\ M}$ (kcal mol^{-1})	From kinetics† ($\Delta\Delta G^{\ddagger-D}_{int} - \Delta\Delta G^{\ddagger-N}_{int\ 3\ M}$) (kcal mol^{-1})
Ala16–Leu49	−0.8 ± 0.4	−0.2 ± 0.1
Ala16–Ile57	−2.0 ± 0.2	−1.1 ± 0.2

*Calculated from the $\Delta\Delta G^{D-N}_{int\ 3\ M}$ values in Table 5. †For a two-state reaction, the difference ($\Delta\Delta G^{\ddagger-D}_{int} - \Delta\Delta G^{\ddagger-N}_{int}$; from Tables 2 and 4) should be equivalent to $\Delta\Delta G^{D-N}_{int}$. The kinetic measurements for A16G/I57A indicate the presence of strain in the native state, as is found from equilibrium measurements on the protein. Ala16 and Leu49, on the other hand, appear not to interact significantly in the native protein. Negative coupling energies are unfavourable.

homologous proteins. This would suggest a preference for a smaller aliphatic sidechain at position 57 in the packing of the hydrophobic core of the protein.

Conclusions

Mutation of the highly conserved residues 16, 49 and 57 in the folding nucleus of CI2 leads to the most dramatic decreases in stability and folding rates found in over 100 mutants studied so far [17]. Our double-mutant analysis reveals that there is strain in the conserved hydrophobic core of native CI2, which is released upon unfolding. The strain is reduced in the transition state for folding. Thus, key interactions between residues involved in the formation of the folding nucleus are not optimised for the stability of the ground state. The hydrophobic core of CI2 rather shows some 'frustration' between Ala16 and Ile57.

An independent computer simulation predicts that when folding rates for protein-like sequences are optimised, residues in the folding nucleus are highly conserved (V.I. Abkevich, L.A. Mirny and E.I. Shakhnovich, personal communication). Our study suggests that optimal packing in a hydrophobic core may be at the expense of enzyme activity or rate of folding.

Materials and methods

Chemicals

The buffer used in the equilibrium and kinetic folding experiments was 2-(*N*-morpholino)-ethanesulfonic acid (MES) and was purchased from Sigma (St. Louis, MO, USA). Guanidinium chloride was sequanal grade from Pierce Chemicals (Rockford, IL, USA).

Preparation of mutants

Mutagenesis and expression were performed as described previously [17]. Because of the instability of the mutant proteins A16G/L49A and A16G/I57A, a modification of the published purification procedure [36] was used. The pellet obtained by centrifugation after the sonication step was resuspended in 7 M GdmCl. The solution was then dialysed extensively against 5 mM acetate buffer, pH 4.5. Aggregates were removed by centrifugation and the pH of the supernatant was adjusted to 7.8 with 50 mM Tris buffer. The anion-exchange resin, ammonium-sulphate precipitation and gel-filtration chromatography steps which followed were performed as described [36]. The purified proteins were homogenous as judged by SDS-PAGE and ESI-mass spectrometry.

Equilibrium denaturation and kinetic experiments

Equilibrium unfolding experiments were performed at 25°C with a Perkin Elmer LS5B luminescence spectrometer, as described previously [17,36]. Kinetic experiments were performed with an Applied Photophysics Stopped-Flow Spectrophotometer model SX 17MV. Analysis of the equilibrium experiments was performed as described previously [17,36]. Kinetic measurements were performed at 25°C and pH 6.3, as described [17]. The rate of folding, k_f, was measured by a pH jump from acid (20 mM HCl) to a final pH of 6.3. The rate constant for unfolding in water, $k_u^{H_2O}$, was extrapolated from a series of GdmCl jump unfolding experiments using $\ln k_u^{[GdmCl]} = \ln k_u^{H_2O} - m_{\ddagger-N}[GdmCl]$, where $m_{\ddagger-N}$ is a constant proportional to the surface area exposed on unfolding. The $m_{\ddagger-N}$ value for wild-type and mutant proteins are identical within error.

Acknowledgements

A.G.L. is supported by a doctoral fellowship from the Boehringer Ingelheim Fonds (Germany) and L.S.I. by a Beit Memorial Fellowship for Medical Research (UK).

References

1. Jackson, S.E. & Fersht, A.R. (1991). Folding of chymotrypsin inhibitor-2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435.
2. Kuszewski, J., Clore, G.M. & Gronenborn, A.M. (1994). Fast folding of a prototypic polypeptide: the immunoglobulin binding domain of streptococcal protein G. *Protein Sci.* **3**, 1945-1952.
3. Huang, G.S. & Oas, T.G. (1995). Submillisecond folding of monomeric lambda repressor. *Proc. Natl Acad. Sci. USA* **92**, 6878-6882.
4. Schindler, T.S., Herrler, M., Marahiel, M.A. & Schmid, F.X. (1995). Extremely rapid folding in the absence of intermediates. *Nat. Struct. Biol.* **2**, 663-673.
5. Kragelund, B.B., Robinson, C.V., Knudsen, J., Dobson, C.M. & Poulsen, F.M. (1995). Folding of a four-helix bundle: studies of acyl-coenzyme A binding protein. *Biochemistry* **34**, 7217-7224.
6. Fersht, A.R. (1995). Optimization of rates of protein folding: the nucleation-collapse mechanism for the folding of chymotrypsin inhibitor 2 (CI2) and its consequences. *Proc. Natl Acad. Sci. USA* **92**, 10869-10873.
7. Waldburger, C.D., Jonsson, T. & Sauer, R.T. (1996). Barriers to protein-folding – formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA* **93**, 2629-2634.
8. Viguera, A.R., Villegas, V., Avilés, F.X. & Serrano, L. (1997). Favourable native-like helical local interactions can accelerate protein folding. *Fold. Des.* **2**, 23-33.
9. Jacob, M., Schindler, T., Balbach, J. & Schmid, F.X. (1997). Diffusion control in an elementary protein folding reaction. *Proc. Natl Acad. Sci. USA* **94**, 5622-5627.
10. Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L. & Oas, T.G. (1997). The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat. Struct. Biol.* **4**, 305-310.
11. Abkevich, V.I., Gutin, A.M. & Shakhnovich, E.I. (1996). Improved design of stable and fast-folding model proteins. *Fold. Des.* **1**, 221-230.
12. Shakhnovich, E., Abkevich, V. & Ptitsyn, O. (1996). Conserved residues and the mechanism of protein-folding. *Nature* **379**, 96-98.
13. Mirny, L.A., Abkevich, V. & Shakhnovich, E.I. (1996). Universality and diversity of the protein folding scenarios: a comprehensive analysis with the aid of a lattice model. *Fold. Des.* **1**, 103-116.
14. Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z. & Wolynes, P.G. (1996). Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* **1**, 441-450.
15. Wolynes, P.G., Luthey-Schulten, Z. & Onuchic, J.N. (1996). Fast folding experiments and the topography of protein-folding energy landscapes. *Chem. Biol.* **3**, 425-432.
16. Otzen, D.E., Itzhaki, L.S., elMasry, N.F., Jackson, S.E. & Fersht, A.R. (1994). The 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.
17. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods – evidence for a nucleation-condensation mechanism for protein-folding. *J. Mol. Biol.* **254**, 260-288.
18. Itzhaki, L.S., Neira, J.L., Ruiz-Sanz, J., de Prat Gay, G. & Fersht, A.R. (1995). Search for nucleation sites in small fragments of chymotrypsin inhibitor-2. *J. Mol. Biol.* **254**, 289-304.

19. de Prat Gay, G., *et al*., & Fersht, A.R. (1995). Conformational pathway of the polypeptide chain of chymotrypsin inhibitor-2 growing from its N-terminus *in vitro*. *J. Mol. Biol.* **254**, 968-979.
20. Neira, J.L., Itzhaki, L.S., Ladurner, A.G., Davis, B., de Prat Gay, G. & Fersht, A.R. (1997). Following cooperative formation of secondary and tertiary structure in a single protein module. *J. Mol. Biol.* **268**, 185-197.
21. Matouschek, A. & Fersht, A.R. (1991). Protein engineering in analysis of protein folding pathways and stability. *Methods Enzymol.* **202**, 82-112.
22. Daggett, V., Li, A.J., Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. (1996). Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **257**, 430-440.
23. Li, A.J. & Daggett, V. (1996). Identification and characterization of the unfolding transition-state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* **257**, 412-429.
24. Shoemaker, B.A., Wang, J. & Wolynes, P.G. (1997). Structural correlations in protein folding funnels. *Proc. Natl Acad. Sci. USA* **94**, 777-782.
25. Meiering, E.M., Serrano, L. & Fersht, A.R. (1992). Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* **225**, 585-589.
26. Anderson, D.E., Hurley, J.H., Nicholson, H., Baase, W.A. & Matthews, B.W. (1993). Hydrophobic core repacking and aromatic–aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117→Phe. *Protein Sci.* **2**, 1285-1290.
27. Schreiber, G., Buckle, A.M. & Fersht, A.R. (1994). Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* **2**, 945-951.
28. Lim, W.A., Hodel, A., Sauer, R.T. & Richards, F.M. (1994). The crystal structure of a mutant protein with altered but improved hydrophobic core packing. *Proc. Natl Acad. Sci. USA* **91**, 423-427.
29. Lee, K.N., Park, S.D. & Yu, M.H. (1996). Probing the native strain in alpha(1)-antitrypsin. *Nat. Struct. Biol.* **3**, 497-500.
30. Ryu, S.E., Choi, H.J., Kwon, K.S., Lee, K.N. & Yu, M.H. (1996). The native strains in the hydrophobic core and flexible reactive loop of a serine protease inhibitor: crystal structure of an uncleaved α_1 antitrypsin at 2.7 Å. *Structure* **4**, 1181-1192.
31. 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.
32. 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.
33. 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.
34. Horovitz, A., Serrano, L. & Fersht, A.R. (1991). COSMIC analysis of the major alpha-helix of barnase during folding. *J. Mol. Biol.* **219**, 5-9.
35. Wells, J.A. (1990). Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517.
36. Jackson, S.E., Moracci, M., elMasry, N., Johnson, C.M. & Fersht, A.R. (1993). The effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor-2. *Biochemistry* **32**, 11259-11269.
37. Jackson, S.E., elMasry, N. & Fersht, A.R. (1993). 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.
38. Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A.R. (1990). Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**, 1031-1044.
39. Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A.R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **29**, 9343-9352.
40. 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.
41. Kraulis, P. (1991). Molscript, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
42. Merritt, E.A. & Murphy, M.E.P. (1994). Raster 3D Version 2.0 – A program for photorealistic molecular graphics. *Acta Crystallogr. D.* **50**, 869-873.